(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau



and the same

(43) International Publication Date 17 May 2001 (17.05.2001)

PCT

(10) International Publication Number WO 01/35316 A2

(51) International Patent Classification⁷: G06F 19/00

(21) International Application Number: PCT/US00/30863

(22) International Filing Date:

10 November 2000 (10.11.2000)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

09/438,566 10 November 1999 (10.11.1999) US 09/704,362 1 November 2000 (01.11.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

 US
 09/438,566 (CIP)

 Filed on
 10 November 1999 (10.11.1999)

 US
 09/704,362 (CIP)

 Filed on
 1 November 2000 (01.11.2000)

- (71) Applicants (for all designated States except US): STRUCTURAL BIOINFORMATICS, INC. [US/US]; 10929 Technology Place, San Diego, CA 92127 (US). QUEST DIAGNOSTICS INVESTMENTS INCORPORATED [US/US]; 300 Delaware Avenue, Wilmington, DE 19899 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): RAMNARAYAN,

Kalyanaraman [IN/US]; 11674 Springside Road, San Diego, CA 92128 (US). MAGGIO, Edward, T. [US/US]; 18775 Bernardo Trails Drive, San Diego, CA 92128 (US). HESS, P., Patrick [US/US]; 54 Via Zeragoza, San Clemente, CA 92672 (US).

- (74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 4250 Executive Square, Suite 700, La Jolla, CA 92037 (US).
- (81) Designated States (national): AE, AG, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

 Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(\$4) Title: USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS AND CLINICAL APPLICATIONS

(57) Abstract: Provided herein are computer-based methods for generating and using three-dimensional (3-D) structural models of target molecules and databases containing the models. The targets can be protein structural variants derived from genes containing polymorphisms. The models are generated using molecular modeling techniques and are used in structure-based drug design studies for identifying drugs that bind to particular structural variants in structure-based drug design studies, for designing allele-specific drugs and population-specific drugs and for predicting clinical responses in patients. Computer-based methods for predicting drug resistance or sensitivity via computational phenotyping are also provided. Databases containing protein structural variant models are also provided.

-1-

USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS AND CLINICAL APPLICATIONS

RELATED APPLICATIONS

5

10

15

20

25

30

Benefit of priority to the following applications is claimed herein:
U.S. application Serial No. 09/438,566 to Kalyanaraman Ramnarayan,
Edward T. Maggio and P. Patrick Hess, filed November 10, 1999 entitled
"USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF
GENETIC POLYMORPHISMS IN PHARMACOGENOMICS FOR DRUG
DESIGN AND CLINICAL APPLICATIONS"; and U.S. application Serial No.
(Attorney Dkt. No. 24737-1906B) to Kalyanaraman Ramnarayan, Edward
T. Maggio and P. Patrick Hess, filed November 1, 2000, entitled "USE OF
COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC
POLYMORPHISMS IN PHARMACOGENOMICS FOR DRUG DESIGN AND
CLINICAL APPLICATIONS."

Where permitted the above-noted applications are incorporated by reference in their entirety. Also incorporated by reference in its entiretly is U.S. application Serial No. (attorney docket no. 24737-1906C), filed November 10, 2000, to entitled "USE OF COMPUTATIONALLY DERIVED PROTEIN STRUCTURES OF GENETIC POLYMORPHISMS IN PHARMACOGENOMICS AND CLINICAL APPLICATIONS."

Incorporation by reference of Tables provided on Compact Disks

For US purposes and where permitted, an electronic version on compact disk (CD) ROM of Tables 4 and 5, which set forth coordinates for three-dimensional structures of proteins in the database described herein is filed herewith, and, where permitted and for US purposes, the contents thereof is incorporated by reference in its entirety. Table 4 is the HIV reverse transcriptase coordinates, and Table 5 is the HIV protease coordinates. The files that contain Table 4 are entitled 1906TAB.PC1 and 1906TAB.PC2, created on November 10, 2000, and are 59,538 kilobytes

PCT/US00/30863 WO 01/35316

-2-

and 304 kilobytes, respectively, and the file that contains Table 5 is entitled 1906TAB.PC3, created on November 10, 2000, and contains 11,413 kilobytes.

FIELD OF THE INVENTION

~~· 5

10

20

25

The present invention is related to computer-based methods and relational databases that use three-dimensional (3-D) protein structural models derived from genetic polymorphisms in the areas of computerassisted drug design and the prediction of clinical responses in patients.

BACKGROUND OF THE INVENTION

Recent advances in molecular biology, such as the discovery and identification of large numbers of genes and the sequences thereof encoded in the genomes of humans, other mammals and infectious disease agents, have contributed to the identification of a large number of proteins, biological receptors and other macromolecules and complexes 15 that are promising therapeutic targets. Based on the information derived from the gene sequences, the three-dimensional (3-D) molecular structures of the corresponding target proteins or receptors can be determined.

Since 3-D protein structure is related to biological function, structure-based drug design is an increasingly useful methodology that has made a great impact in the design of biologically active lead compounds. Drug designers can design and screen potential new drugs via computational methods, such as docking or binding studies, before actually beginning patient testing. These experiments can be performed in silico at a tiny fraction of the clinical cost.

The resulting molecules, while serving as lead compounds, often have unpredictable effects when employed in clinical trials. In addition, it has been observed that existing drugs with known clinical efficacy far often fail to achieve beneficial results when given to particular patients, or particular subpopulations, such as ethnic groups, of patients. Genetic stratification of a population can be the difference between drug failure and drug approval. Hence there is a need to develop methods to improve the drug discovery process. Therefore, it is an object herein to provide, among a variety of benefits, methods and products that address and solve these problems. In particular, it is an object herein to provide computationally-based methods for drug design, clinical testing protocols, identification of new drug candidates and drug therapies; for predicting drug sensitivity and resistance and other methods.

10 SUMMARY OF THE INVENTION

15

20

25

Provided herein are computer-based methods for generating and using three-dimensional (3-D) structural models of target biomolecules, particularly polymorphic and allelic variants. Also provided herein are databases that contain the sequences of such variants and also the 3-D structure of the variants for use with the methods.

Genetic polymorphisms arise, for example, as a result of gene sequence differences or as a result of post-translational modifications, including glycosylation. Hence genetic polymorphisms are manifested as gene products and proteins having variant structures. The variant structures result in differences in biological responses among the originating organisms. These differences in response, include, but are not limited to, differences among patient responses to a particular drug, effective dosage differences, and side effects. With respect to infectious organisms, some polymorphisms may arise that convey resistance or susceptibility to particular drug therapies by the altering the drug target structure.

Structural changes that arise as a result of genetic polymorphisms are not of unlimited variety, since 3-D structure impacts upon function. A knowledge of the repertoire of the fine differences among generally similar

-4-

3-D structures of particular proteins will permit design of drugs that bind to the most polymorphisms, drugs that induce the fewest side-effects, and drugs that are more effective against infectious agents. Knowledge of these structures ultimately will permit patient-specific or subpopulation-specific, such as ethic, age, or gender groups, design or selection of drugs.

The methods that are provided are for determining and using 3dimensional (3-D) protein structures that are derived from genetic polymorphisms to understand differences in biological activity that result from the polymorphisms, and to use this understanding to aid in the identification of potential new drug candidates and drug therapies. Also provided are methods for analyzing 3-D structures of protein structural variant targets derived from genetic polymorphisms to identify common structural features among the variants; methods for identifying structural changes in target proteins that are associated with multiple mutations arising from genetic polymorphisms and correlating this information with biological activity; methods for using clinical data in conjunction with structural variants derived from genetic polymorphisms to understand and predict the pharmacological effects and clinical outcomes for drugs or potential drugs. Also provided are methods for generating 3-D protein structures derived from a given genotype to analyze protein-drug binding in silico to predict drug sensitivity or resistance. Also provided are databases that are used in methods provided herein and methods for generating the databases.

10

15

20

25

In particular, target biomolecules are protein structural variants encoded by genes containing genetic variations, or polymorphisms. 3-D models of the structures of proteins are determined. The models are generated using molecular modeling techniques, such as homology modeling. The resulting models are then used in the methods provided

-5-

herein, which include structure-based drug design studies to design and identify drugs that bind to particular structural variants; structure-based drug design studies and to predict clinical responses in patients; and to design drugs that bind to all or a substantial portion of allelic variants of a target, to thereby increase the population of patients for whom a particular drug will be effective and/or to decrease the undesirable side-effects in a larger population.

Hence, computer-based methods of drug design based on target protein structural models derived from genetic polymorphisms are provided. The methods involve obtaining one, preferably two or more amino acid sequences of a target protein that is the product of a gene exhibiting genetic polymorphisms, where sequences represent different genetic polymorphisms, and generating 3-D protein structural variant models from the sequences. Structure-based drug design techniques are used to design potential new drug candidates or to suggest modifications to existing drugs based on predicted intermolecular interactions of the drugs or drug candidates with the models. Alternatively, drug molecules can be computationally docked with 3-D protein structural variant models based upon the sequences and energetically refined before performing structure-based drug design studies.

10

15

20

25

In preferred embodiments, binding interactions between a drug or potential new drug candidate molecules and the structural variants are calculated in order to optimize intermolecular interactions between drug or potential drug molecules and the structural variant models or to select drug therapies for patients by determining a drug or drugs that have favorable binding interactions with the structural variant models.

In other embodiments, the binding interactions are determined by calculating the free energy of binding between the protein structural variant model and a docked molecule; and decomposing the total free

-6-

energy of binding based on the interacting residues in the protein active site.

~"· 5

10

15

20

25

After the protein structural variant models are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models. The conserved structural features can serve as scaffolds or pharmacophore models into which potential drugs or modified drugs are docked. For example, the selected model structures may represent the structural variants resulting from the most commonly occurring genetic polymorphisms or from genetic polymorphisms found in a specific patient subpopulation, such as a particular age group, ethnic or racial group, sex, or other subpopulation. Alternatively, the models may be selected based on clinical information, for example, the structural variants may be derived based on patients receiving a specific treatment regimen or exhibiting a particular clinical response to a given drug or on the duration of a particular drug treatment.

The methods provided herein can be used for predicting clinical responses in patients based on genetic polymorphisms. For example, a structural variant model derived from a subject, such as a human patient, exhibiting a particular genetic polymorphism is generated and screened against a number of reference protein structural variant models derived from genetic polymorphisms of the same gene in other such subjects. In certain embodiments, the reference structures are stored in a database, preferably with observed clinical data associated with the structures, or polymorphisms. The structural variant model from the subject is compared to a reference structures, for example, by database searching, in order to identify reference structural variants that are similar to the model structure derived from the subject. Based on the premise that structurally similar targets will have similar clinical responses, a clinical outcome can be predicted for the patient based on the structures

5

10

15

20

25

identified through structural comparison or database searching. This information can also be used in the design and analysis of clinical trials; it can also be used for selecting appropriate therapies for a subject in instances in which the subject is a patient and the protein is a drug target.

The methods are also used to design therapeutic agents that are active against biological targets that have become drug resistant, particularly due to genetic mutations. In certain embodiments, 3-D protein structural variant models are generated for a target protein in which genetic mutations have occurred and against which a given drug is no longer biologically active. The models are compared to 3-D protein structural variant models of the target protein against which the drug has biological activity in order to identify structural differences between the susceptible and resistant targets. The differences can be used to understand the structural contributions to drug resistance, and this information can be utilized in structure-based drug design calculations to identify new drugs or modifications to the existing drug that circumvent the resistance problem.

A computer-based method for identifying compensatory mutations in a target protein is also provided. The method involves obtaining the amino acid sequence of a target protein containing multiple amino acid mutations that is expressed in a patient, where the structure of a form of the target protein that responds to a particular drug, including the active site, has been structurally characterized; generating a 3-D structural model of the mutated protein; comparing the structure of the mutated protein with the form of the protein that responds to the drug to identify structural differences and/or similarities arising from the mutations; comparing the biological activities of the drug against the mutated protein and the form of the protein that responds to the drug to determine the

-8-

effects of the mutations on drug response; and identifying the mutations in the protein that affect biological activity based on the comparisons. The target biolmolecules can also be used in a method referred to herein as computational phenotyping to predict drug sensitivity or resistance for a given genotype. These computer-based method for identifying phenotypes in silico are provided. The methods involve obtaining from a patient/specimen, such as a body fluid or tissue sample, including blood, cerebral spinal fluid, urine, saliva, sweat and tissue samples, the amino acid sequence of a target protein; generating a 3-D structural model of the target protein; performing protein-drug binding analyses; and predicting drug sensitivity or resistance based on the protein-drug binding analyses.

Molecular structure databases containing protein structural variant models produced by the methods are also provided. The databases may also contain biological or clinical data associated with the structural variants. The databases can be interfaced to a molecular graphics package for visualization and analysis of the 3-D molecular structural models. In particular, databases containing the 3-D structures of polymorphic variants of selected target genes, particularly pharmaceutically significant genes with pharmaceutically significant gene products, such as proteases and polymerases, including reverse transcriptases, and receptors, such as cell surface receptors, are provided. The databases may be stored an provided on any suitable medium, including, but are not limited to, floppy disks, hard drives, CD-ROMS and DVDs.

15

20

25

Also provided are relational databases for managing and using information relating to genetic polymorphisms. The databases contain 3-D molecular coordinates for structural variants derived from genetic polymorphism, a molecular graphics interface for 3-D molecular structure

15

25

visualization, computer functionality for protein sequence and structural analyses and database searching tools. The databases may further include observed clinical data associated with the genetic polymorphism. The databases provide a means to design the allele-specific drugs and also to identify among alleles common or conserved structural features that can serve as the target for drug design.

The databases can also be used for identification of invariant residues and regions of a target biomoleucle, such as an HIV protease or reverse transcriptase. The identified invariant regions are then used to computationally screen compounds, preferably small molecules by assessing binding interactions. The compounds so-identified serve as candidates for drugs that will be effective for a larger proporation of a population or against a broader range of variants of a pathogen, where the target protein is from a pathogens.

Systems, including computers, containing the databases also are provided herein. Any computer known to those of skill in the art for maintaining such databases is contemplated. User interfaces for accessing and manipulating the databases and content thereof are also provided.

20 BRIEF DESCRIPTION OF THE DRAWINGS

- FIG. 1 illustrates a method for creating a protein structural variant relational database.
- FIG. 2 is a flow chart that describes one method used to generate structural variant models derived from genetic polymorphisms and to use the models in structure-based drug design studies.
- FIG. 3 is a flow chart that describes an alternative method used to generate structural variant models derived from genetic polymorphisms and to use the models in structure-based drug design studies.

- FIG. 4 shows the correlation between experimental and calculated changes of binding energy upon ligand modifications in the binding site of NS3.
- FIG. 5 shows a comparison of calculated *versus* experimental binding free energy changes for complexes of the tumor necrosis factor (TNF) receptor with different inhibitors.
 - FIG. 6 shows the HIV PR inhibitors approved by the FDA.
 - FIG. 7 shows the frequency versus amino acid residue plot of HIV PR.
- 10 FIG. 8 shows frequency analysis of 10591 HIV PR Sequences, where ResNum is the residue number; TotOcc is the total occurrence of the mutation; Dist is the distance of the mutating residue from approximate center of active site (Asp28); WtAA is the amino acid in the wild type protein; NumMut is the number of mutations; and MutList is a list of amino acid mutations.
 - FIG. 9 is a block diagram of an exemplary computer.
 - FIG. 10 is a graphical representation of a relational database.
 - FIG. 11 is a tabulation of the 3-D coordinates of a representative entry in a database that includes 3-D structures.

20 DETAILED DESCRIPTION OF THE INVENTION

A. Definitions

30

- B. Computer-based methods of drug design based on genetic polymorphisms
- 25 1. Methods for obtaining amino acid sequences of a target protein
 - 2. Generation of 3-D protein structural variant models
 - a. Homology Modeling
 - b. Ab initio generation of 3-D structures
 - c. Crystal structures
 - 3. Use of 3-D structural variant models in drug design
 - a. Selection of relevant structural variants
 - b. Drug design
 - c. Computational docking

-11-

d. Free energy of binding studies

C. Applications of computer-based methods

5 1. Genetic polymorphisms and structure-based drug design

2. Drug resistance

3. Identification of conserved structural features or pharmacophores

4. Identification of compensatory structural changes

10 5. Clinical Applications

D. Creation of 3-D Structural Polymorphism Databases

1. Exemplary Databases and generation thereof

2. Computer systems and Database

E. Computational phenotyping

A. Definitions

15

30

35

Unless defined otherwise, all technical and scientific terms used
herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents, patent applications, published patent applications and publications referred to herein are, unless noted otherwise, incorporated by reference in their entirety. In the event a definition in this section is not consistent with definitions elsewhere, the definition set forth in this section will control.

As used herein, polymorphism refers to a variation in the sequence of a gene in the genome amongst a population, such as allelic variations and other variations that arise or are observed. Genetic polymorphisms refers to the variant forms of gene sequences that can arise as a result of nucleotide base pair differences, alternative mRNA splicing or post-translational modifications, including, for example, glycosylation. Thus, a polymorphism refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. These differences can occur in coding and non-coding portions of the genome, and can be manifested or detected as differences in nucleic acid

-12-

sequences, gene expression, including, for example transcription, processing, translation, transport, protein processing, trafficking, DNA synthesis, expressed proteins, other gene products or products of biochemical pathways or in post-translational modifications and any other differences manifested among members of a population. A single nucleotide polymorphism (SNP) refers to a polymorphism that arises as the result of a single base change, such as an insertion, deletion or change in a base.

5

10

15

20

25

A polymorphic marker or site is the locus at which divergence occurs. Such site may be as small as one base pair (an SNP). Polymorphic markers include, but are not limited to, restriction fragment length polymorphisms, variable number of tandem repeats (VNTR's), hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats and other repeating patterns, simple sequence repeats and insertional elements, such as Alu. Polymorphic forms also are manifested as different mendelian alleles for a gene. Polymorphisms may be observed by differences in proteins, protein modifications, RNA expression modification, DNA and RNA methylation, regulatory factors that alter gene expression and DNA replication, and any other manifestation of alterations in genomic nucleic acid or organelle nucleic acids.

As used herein, structural variants proteins refer the variety of 3-D molecular structures or models thereof that result from the polymorphisms. These variants typically arise from transcription and translation of genes containing genetic polymorphisms, but also include diffentially glyocsylated or otherwise post-translationally modified variants that potentially exhibit differential interactions with drugs and drug candidates.

As used herein, binding interactions refer to atomic or physical interactions between molecules including, but not limited to binding free energy, hydrophobic interactions, electrostatic interactions, steric interactions and other interactions that are commonly considered by those of skill in the art to determine the affinity of one molecule to bind to another. Favorable binding interactions refer to binding interactions that promote physical or chemical associations between molecules.

As used herein, a target protein is defined as a protein that is a receptor with which drugs or other ligands, such as small molecule or peptide agonists or antagonists or other proteins or biomacromolecules, such as DNA or RNA, interact to bring about a biological response.

10

15

25

As used herein, structure-based drug design refers to computer-based methods in which 3-D coordinates for molecular structures are used to identify potential drugs that can interact with a biological receptor. Examples of such methods include, but are not limited to, searching of small molecule libraries or databases, conformational searching of a ligand within an active site of identify biologically active conformations or computational docking methods.

As used herein, pharmacogenomics refers to study of the variablity of patient responses to drugs due to inherent genetic differences.

As used herein, computational docking refers to techniques wherein molecules, for example, a ligand and receptor or active site, are fitted together based on complementary interactions, for example, steric, hydrophobic or electrostatic interactions.

As used herein, energetic refinement refers to the use of molecular mechanics simulation techniques, such as energy minimization or molecular dynamics, or other techniques, such as quantum-based approaches, to "adjust" the coordinates of a molecular structural model to bring it into a stable, low energy, conformation. In molecular mechanics

-14-

simulations, the potential energy of a molecular system is represented as a function of its atomic coordinates along with a set of atomic parameters, called a forcefield. Energy minimization refers to a method wherein the coordinates of a molecular conformation are adjusted according to a target function to result in a lower energy conformation. Molecular dynamics refers to methods for simulating molecular motion by inputting kinetic energy into the molecular system corresponding to a specified temperature, and integrating the classical equations of motion for the molecular system. During a molecular dynamics simulation, a system undergoes conformational changes so that different parts of its accessible phase space are explored.

As used herein, clinical data refers to information obtained from patients pertaining to pharmacological responses of the patient to a given drug, including, but not limited to efficacy data, side effects, resistance or susceptibility to drug therapy, pharmacokinetics or clinical trial results.

As used herein, patient histories, include medical histories and other any information, such as parental medical histories, dates and places of birth of the patient and parents, number of siblings, number of children and other such data.

As used herein, compensatory mutations are mutations that act in concert with active site mutations by compensating for functional deficits caused by changes or mutations that affect binding in the active site.

20

25

As used herein, a relational database is a collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables. Such databases are readily available commercially, for example, from Oracle, IBM, Microsoft, Sybase, Computer Associates, SAP, or multiple other vendors.

As used herein, a phenotype refers to a set of parameters that includes any distinguishable trait of an organism. A phenotype can be physical traits and can be, in instances in which the subject is an animal, a mental trait, such as emotional traits. Some phenotypes can be determined by observation elicited by questionnaires or by referring to prior medical and other records. For purposes herein, a phenotype is a parameter around which the database can be sorted.

As used herein, genotype refers to a specific gene or totality of genetic information in a specific cell or organism.

10

15

20

25

As used herein, haplotype refers refers to two or more polymorphism located on a single DNA strand. Hence, haplotyping refers to identification of two or more polymorphisms on a single DNA strand. Haplotypes can be indicative of a phenotype.

As used herein, a parameter is any input data that will serve as a basis for sorting the database. These parameters will include phenotypic traits, medical histories, family histories and any other such information elicited from a subject or observed about the subject. A parameter may describe the subject, some historical or current environmental or social influence experienced by the subject, or a condition or environmental influence on someone related to the subject. Paramaters include, but are not limited to, any of those described herein, and known to those of skill in the art.

As used herein, computational phenotyping, refers to computer-based processes that assess the phenotype resulting from a particular genotype. The phenotype describes observables, such as, but are not limited to, the structure of the encoded protein, its functional morphological and structural attributes. In particular, as contemplated herein, the phenotype that is assessed is the interaction of a protein with a particular compounds, particularly a drug. As exemplified herein, the

-16-

method provides a means to select an effective drug for a particular subjects, particularly mammals, or class thereof.

As used herein, a database refers to a collection of data; in this case data relating to polymorphic variants. Hence a database contains the nucleic acid sequences encoding the variants, or a portion of the variant, such as a portion contianing the active site or targetted site. Additionally, the database may contain other information related to each entry, including but are not limited to, the corresponding 3-D structure of the encoded protein (or a portion thereof) and information regaring the source of each sequence. Some of the entries in a database may be identical, and for purposes herein, a database contains at least 2 different entries, typically far more than 2 entries. The number of entries depends upon the protein of interest and variety and number of polymorphisms that exist. Generally a database will have at least 10 different entries, typically more than 100, more than 500, more than 1000, more than 2000, 3000, 4000, 5000, 8000, 10,000, 50,000, 100,000 and greater. Databases herein containing 20,000 entries and more have been generated and are exemplified herein.

10

15

20

25

As used herein, a relational database stores information in a form representative of matrices, such as two-dimensional tables, including rows and columns of data, or higher dimensional matrices. For example, in one embodiment, the relational database has separate tables each with a parameter. The tables are linked with a record number, which also acts as an index. The database can be searched or sorted by using data in the tables and is stored in any suitable storage medium, such as floppy disk, CD rom disk, hard drive or other suitable medium.

As used herein, a profile refers to information relating to, but not limited to and not necessarily including all of, age, sex, ethnicity, disease

-17-

history, family history, phenotypic characteristics, such as height and weight and other relevant parameters.

As used herein, a biopolymer includes, but is not limited to, nucleic acid, proteins, polysaccharides, lipids and other macromolecules. Nucleic acids include DNA, RNA, and fragments thereof. Nucleic acids may be derived from genomic DNA, RNA, mitochondrial nucleic acid, chloroplast nucleic acid and other organelles with separate genetic material.

As used herein, a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence. By the term "substantially homologous" is meant having at least 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

As used herein, a receptor refers to a molecule that has an affinity for a given ligand. Receptors may be naturally-occurring or synthetic molecules. Receptors may also be referred to in the art as anti-ligands. As used herein, the terms, receptor and anti-ligand are interchangeable. Receptors can be used in their unaltered state or as aggregates with other species. Receptors may be attached, covalently or noncovalently, or in physical contact with, to a binding member, either directly or indirectly via a specific binding substance or linker. Examples of receptors, include, but are not limited to: antibodies, cell membrane receptors surface receptors and internalizing receptors, monoclonal antibodies and antisera reactive with specific antigenic determinants (such as on viruses, cells, or other materials), drugs, polynucleotides, nucleic acids, peptides, cofactors, lectins, sugars, polysaccharides, cells, cellular membranes, and organelles.

20

25

Examples of receptors and applications using such receptors, include but are not restricted to:

a) enzymes: specific transport proteins or enzymes essential to survival of microorganisms, which could serve as targets for antibiotic (ligand) selection;

·· 5

15

25

- b) antibodies: identification of a ligand-binding site on the antibody molecule that combines with the epitope of an antigen of interest may be investigated; determination of a sequence that mimics an antigenic epitope may lead to the development of vaccines of which the immunogen is based on one or more of such sequences or lead to the development of related diagnostic agents or compounds useful in therapeutic treatments such as for auto-immune diseases:
- c) nucleic acids: identification of ligand, such as protein or RNA, binding sites;
- d) catalytic polypeptides: polymers, preferably polypeptides, that are capable of promoting a chemical reaction involving the conversion of one or more reactants to one or more products; such polypeptides generally include a binding site specific for at least one reactant or reaction intermediate and an active functionality proximate to the binding site, in which the functionality is capable of chemically modifying the 20 bound reactant (see, e.g., U.S. Patent No. 5,215,899);
 - e) hormone receptors: determination of the ligands that bind with high affinity to a receptor is useful in the development of hormone replacement therapies; for example, identification of ligands that bind to such receptors may lead to the development of drugs to control blood pressure; and
 - f) opiate receptors: determination of ligands that bind to the opiate receptors in the brain is useful in the development of less-addictive replacements for morphine and related drugs.

As used herein, prion refers to an infectious pathogen that causes central nervous system spongiform encephalopathies in humans and animals. No nucleic acid component is necessary for the infectivity of prion protein (see, e.g., U.S. Patent No. 5,808,969).

- 5

10

15

20

25

As used herein, a ligand is a molecule that is specifically recognized by a particular receptor. Examples of ligands, include, but are not limited to, agonists and antagonists for cell membrane receptors, toxins and venoms, viral epitopes, hormones (e.g., steroids), hormone receptors, opiates, peptides, enzymes, enzyme substrates, cofactors, drugs, lectins, sugars, oligonucleotides, nucleic acids, oligosaccharides, proteins, and monoclonal antibodies.

As used herein, complementary refers to the topological compatibility or matching together of interacting surfaces of a ligand molecule and its receptor. Thus, the receptor and its ligand can be described as complementary, and furthermore, the contact surface characteristics are complementary to each other.

As used herein, a ligand-receptor pair or complex formed when two macromolecules have combined through molecular recognition to form a complex.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be determined, for example, by comparing sequence information using a GAP computer program. The GAP program utilizes the alignment method of Needleman and Wunsch (*J. Mol. Biol.* 48:443 (1970), as revised by Smith and Waterman (*Adv. Appl. Math.* 2:482 (1981). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e., nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing

a value of 1 for identities and 0 for non-identities) and the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745 (1986), as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps.

Whether any two nucleic acid molecules have nucleotide sequences that are at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA 85*:2444 (1988). Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine identity

10

15 In general, sequences are aligned so that the highest order match is obtained. "Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g.: Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., 20 Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number 25 of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., SIAM J Applied Math 48:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to

-21-

Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., SIAM J Applied Math 48:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., Nucleic Acids Research 12(I):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., et al., J Molec Biol 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide.

10

15

20

25

As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to a reference polypeptide. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

As used herein, AMBER is a force field well known in the arts and designed for the study of proteins and nucleic acids as defined in Weiner et al. J. Comput. Chem. (1986) 7:230-252, where a modified AMBER (version 3.3) force field is a fully vectorized version of AMBER (version

3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy. AMBER is available in commercially available molecular modeling programs such as, but not limited to, Macromodel (Columbia University).

As used herein, ECEPP (Empirical Conformational Energies of Peptides Program) is a force field well know in the arts (US Patent No. 5,910,478; 5,846,763). ECEPP/3 refers to version 3 of this well known force field.

As used herein, QSAR refers to structure-activity relationship.

10 As used herein, vdw refers to van der Waals.

5

15

20

30

As used herein, RMSD refers to root mean-squared deviation.

As used herein, medical history refers to the parameters and data typically obtained by a physician when examining a subject or other such professional when examining other mammals, and includes such information as prior diseases, age, weight, height, sex and other information. For purposes, the subjects that serve as the source of the samples from which nucleic acids encoding polymorphisms are isolated, include animals, plants, pathogens and any organism that has nucleic acid that exhibits polymorphism. In this context medical history refers to information pertinent to the particular organism.

As used herein, subject history, refers to data such as locale in which the subject was born, raised or resident or visited, and parental history and other such information.

As used herein, a drug is an agent that binds to or interacts with a targeted protein. For purposes, a therapeutic agent is a drug.

B. Computer-based methods of drug design based on genetic polymorphisms

Methods for computer-based drug design based on genetic polymorphisms are provided. The methods includes the steps of obtaining one or more, preferably two or more, amino acid sequences of a target

-23-

protein that is the product of a gene exhibiting genetic polymorphisms; generating 3-dimensional (3-D) protein structural variant models of all or a portion of the protein from the sequences; and based upon the structures of the 3-D models, designing drug candidates or modifying existing drugs based on the predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants or portions thereof by computationally docking drug molecules with the target protein models; and then, optionally energetically refining the docked complexes; determining the binding interactions between the drug or potential new drug candidate molecules and the models by calculating the free energy of binding of the docked complexes and decomposing the total free energy of binding based on interacting residues in the protein active site or sites deemed important for protein activity.

10

15

20

25

A variety of methods that include these steps are provided. Such methods have particularl application, for example, in predicting patient responses. As noted, patients exhibit variable responses to drugs. For some patients a drug may be very beneficial and achieve a desired response; whereas for other patients, with the same disorder, the same drug will have little or no effect. It is known that individuals as well as groups of individuals exhibit a variety of genetic polymorphisms. As described herein, the presence or absence of such polymorphisms can be correlated with the variability of patient responses to drugs.

It is shown herein that by understanding how genetic polymorphisms affect 3-D protein structure of a drug target, for example, it is possible to ascertain the interaction of a particular drug with the target in a particular patient or groups of patients. Based upon this interaction, the outcome can be predicted. It will be possible to determine whether a patient will benefit from a drug or be at risk for a particular side effect. It is possible to predict these responses before exposure to the drug. These

methods also permit rational design of drugs that can treat various populations or ultimately even individuals. These differences and effects can also be taken into account to design drugs that are not dependent upon a particular polymorphism.

Hence, the knowledge derived from understanding the effects of genetic polymorphisms can be used to develop and apply therapeutics more effectively, make clinical trials more successful, for example, by permitting selection of test subjects with the same polymorphism or with polymorphisms for which the drug is designed to interact effectively.

-- 5

10

20

25

It is shown herein that it is advantageous to use 3-D molecular structures in drug design rather than to consider primary sequence alone. For example, most drugs target proteins either in the afflicted organism or in a pathogen. Disease, drug action and toxicity are all manifested at the protein level. Although the nucleotide sequences of genetic polymorphisms might appear to be quite different, the resulting protein targets may have similar shapes and, therefore, the protein biological function might be the same. Conversely, although genetic polymorphism sequences might appear similar, the resulting proteins may have critical differences in their 3-D structures that greatly affect biological activity. Thus, use of 3-D protein structure models in such methods provide advantages not heretofor realized. Methods for generating 3-D structures

Once the protein target structural models have been selected, structure-based drug discovery methodologies, for example, computational screening or docking programs and methods (e.g., DOCK (available from University of Ca, San Francisco; and AUTODOCK available from Scripps Research Institute, La Jolla), are used to design biologically-active compounds based on the 3-D structures of the biomolecular receptors. Using these methods, drug designers can identify and

are known to those of skill in the art and are also provided herein.

-25-

computationally rank the various potential clinical drug candidates for maximum efficacy, thereby performing drug discovery *in silico* and avoiding the tedious time and expense associated with *in vitro* drug discovery methods.

5

10

15

20

In addition to drug design applications, the information derived from studying the structures of biological targets can be used to understand and predict biological responses in patients, such as efficacy, toxicity, drug resistance and other pharmacological effects. Since human clinical trials may cost upwards of \$100-300 million, it is desirable to predict the outcome to the greatest extent possible for each prospective drug candidate so that the best prospective drug candidates are advanced to clinical trials. As described below, methods are provided herein for selecting populations for clinical trials.

Methods for obtaining amino acid sequences of a target protein

Any protein or gene or encoded mRNA that exhibits polymorphisms, herein referred to as the target protein, in structure is contemplated for use herein and for generating the databases as provided herein. The target protein is a protein, polypeptide, or oligopeptide that includes, but is not limited to, receptors, enzymes, hormones, prions, or any such compound with which drugs or other ligands, such as small molecules, peptide agonists, peptide antagonists, other proteins, nucleic acids and other biomacromolecules, interact to bring about a biological response. These target proteins occur in any organism, including plants and animals, eukaryotes and prokaryotes, including pathogens, such as protozoans, parasites, viruses, includind DNA and retroviruses, and bacteria. The protein or gene can be one expressed in the organism, such as molecule targeted for drug interaction, or one expressed in a pathogen.

The target gene is one that exhibits polymorphisms (i.e., sequence variations among a population) and the target protein is the product of a gene exhibiting genetic polymorphisms, or sequence variations, as described herein. Any gene or protein that exhibits polymorphisms is contemplated herein. In particular, genes that encode proteins, polypeptides, or oligopeptides that are targets for drug interaction are contemplated herein. The genetic polymorphisms can occur in the genes of pathogens (e.g. viruses, bacteriae, and fungi), parasites, plants, animals, and humans. As such, the sequence a target protein can be obtained by the isolation and analysis of the gene or gene product in samples taken from pathogens, parasites, plants, animals, and humans, most preferably from humans.

10

20

25

The genes or proteins may be isolated from any source, such as animal or plant specimens, or the sequences obtained from any source, including known databases. If starting with gene sequences that include single or multiple nucleotide polymorphisms, the amino acid sequences of the translated proteins can be determined. Protein isolation and sequencing methods are well known to those of skill in the art. Alternatively, samples of the target protein can be obtained and sequenced directly from specimens. Multiple sequence analyses can be performed to determine the exact amino acid variations or mutations resulting from the genetic polymorphisms.

Amino acid sequences of target proteins can also be obtained from data banks and databases (e.g. GenBank, Swiss Prot, PIR) and from publications and other sources in which numerous polymorphisms have been identified and mapped. Samples may be obtained from, for example blood and tissue banks, nucleic acid isolated, genes selected or identified and polymorphims can be mapped from such samples.

2. Generation of 3-D protein structural variant models

After the amino acid sequences of target proteins are obtained via the means described in section 1, the 3-D structural models of the sequences of native proteins or of the protein structural variants are then determined. They can be determined through experimental methods, such as x-ray crystallography and NMR, and from structure databases, such as the Protein Databank (PDB). Moreover, 3-D structural models can be determined by using any of a number of well known techniques for predicting protein structures from primary sequences (e.g. SYBYL (Tripos Associated, St. Louis, Mo.), de novo protein structure design programs 10 (e.g. MODELER (MSI, Inc., San Diego, CA) and MOE (Chemical Computing Group, Montreal Canada) and ab initio methods, see, e.g., U.S. Patent Nos. 5,331,573, 5,579,250 and 5,612,895), homology modeling, and ab initio computational analysis. Homology modeling, structure determination based upon x-ray crystallographic structures, and 15 ab initio techniques and combinations of these methods are among those preferred herein.

a. Homology Modeling

evolutionary origin, function and folding patterns. Proteins of related origin and function have conserved sequences and structural features among the members of a homologous family. Using these relationships, a three-dimensional structural model for a protein of unknown structure can be constructed by using composite parts of related proteins in the same family. Where only the primary amino acid sequence of a target protein is known, the sequence can be compared to the sequences of related proteins with known structures (reference proteins), and a model can be built by incorporating the structural attributes of the reference protein together with the sequence of the target protein.

-28-

Sequence homology calculations generally require: the amino acid sequence of the target protein; a high resolution structure for at least one, but preferably more, related reference proteins; and any other related amino acid sequences. The reference proteins include structures which are similar to the target protein, either by sequence, fold, function, or which are polymorphisms of the target protein. The more related protein structures and sequences that are available or determined, the more reliable the technique will be at providing an accurate model.

In constructing a protein model using homology modeling, sequence alignment is performed between the target sequence and any known structures within the protein family. Sequence alignment requires determining the similarity between protein sequences by maximizing the number of matches between the sequences while introducing the minimum number of insertions and deletions. Sequence alignment algorithms are well known in the art, and standard gap penalties (*i.e.*, programs that automatically introduce gaps to maximize alignment and then adjust the percentage of identity by applying penalties for gap number and gap length) and other parameters can be selected by the skilled artisan. Additionally, the 3-D structures of the known reference proteins, preferably, are aligned to give the best overall fit for the proteins in the family. This provides indication of structurally-conserved regions, such as regions of the proteins that do not contain insertions or deletions, among the reference structures.

10

20

25

Once the sequences are aligned and the structurally-conserved regions are identified, the coordinates of the reference proteins can be used to construct a 3-D model of the target structure. Coordinates from the protein backbone of the reference proteins are then used to construct the backbone framework for the target protein structure. Side chains can be constructed, for example, by using side chain coordinates from the

PCT/US00/30863 WO 01/35316

-29-

reference proteins, searching from a database to obtain side chain conformations that fit in with the existing structural framework or by generating side chains ab initio to establish energetically favorable side chain conformations.

5

10

15

25

The non-conserved regions of the unknown protein can be constructed, for example, using database searching. A database of known protein structures (e.g., PDB) can be searched to identify variable regions in other proteins that have a high degree of sequence similarity to the target sequence and that fit onto the existing structural framework of the protein model. Algorithms for performing sequence similarity matching and homology model building are well known in the art and are available commercially (available from Molecular Simulations, Inc., Tripos, Inc. and from numerous academic sources).

The variable regions can also be modeled by fitting the target sequence to a peptide backbone generated by varying phi and psi angles (e.g., by calculating Ramachandran or Balasubramanian plots, see, Balasubramanian (1974) "New type of representation for Mapping Chain Folding in Protein Molecules," Nature 266:856-857) or Balaji plots, see, U.S. Patent Nos. 5,331,573, 5,579,250 and 5,612,895) of the amino acids to give a loop structure that can be integrated into the model 20 structure based on a sterically and energetically reasonable fit (Figure 1).

In a Balasubramanian plot, the peptide is depicted as a series of different vertical lines, each having solid dots and open circles aligned with the corresponding ϕ , ψ angle values on the vertical axis, and where each line corresponds to the particular number of the residue having the plotted ϕ , ψ angles as indicated on a horizontal axis. In the Balaji plot, the values of the ϕ , ψ angles are shown as the base and tip of a vertical wedge (assuming a vertical angular axis), respectively, with a separate wedge being horizontally positioned on the plot as a function of the

5

10

15

20

residue number of the ϕ , ψ angles plotted. The Balaji plot replaces the solid dots and open circles of the Balasubramanian Plot with the base of a wedge and the tip of a wedge, respectively; and further replaces the vertical line joining the dots and open circles of the Balasubramanian plot with the body of the wedge.

b. Ab initio generation of 3-D structures

Alternatively, ab initio methods can be used in combination with an existing partial homologous structure to generate unresolved portions of the target structure. Such methods are described, for example, in U.S. Patent Nos. 5,331,573, 5,579,250 and 5,612,895, which as all patents, applications and publications referenced herein, are each incorporated in their entirety. These methods involve: simulating a real-size primary structure of a polypeptide in a solvent box, i.e., an aqueous environment; shrinking the size of the peptide isobarically and isothermally; and expanding the peptide to its real size in selected time periods, while measuring the energy state and coordinates, i.e., the bonds, angles and torsions of the expanding molecule. As the peptide expands to its full size, it assumes a stable tertiary structure. In most cases, due to the manner in which the expansion occurs, this tertiary structure will be either the most probable structure (i.e., it will represent a global minimum for the structure) or one of the most probable structures. The energy equations used to perform the ab initio simulation are based on the potential energy of the simulated molecule as described using molecular mechanics.

Once a model is built, it can be refined using energy minimization, molecular dynamics calculations, or simulated annealing as described herein. The steric and energetic quality of the structural models is then evaluated by analyzing the structural attributes of the model, such as phi and psi angles (e.g., by calculating Ramachandran or Balasubramanian or

5

10

15

20

25

Balaji plots), or the energetics of the model, such as by calculating energy per residue or strain energy. If the overall quality of the model is not satisfactory, further iterative energy refinement can be performed until the model is considered to be acceptable (i.e., $e_{av} < 1.5$, see below).

A preferred method for generating and refining the structural variant models is illustrated in **FIG. 1**. First, at block 100 of FIG. 1, protein sequence information, derived genetic polymorphisms, is obtained from the methods described earlier. At block 102, the pretein is assigned to a protein superfamily in order to identify related proteins to be used as templates to construct a 3-D model of the protein. If the superfamily is not known, sequence analysis or structural similarity searches can be performed to identify related proteins for use as templates in homology modeling studies, as described herein, as indicated at block 104.

Once the conserved regions of the model are assembled, ab initio loop prediction (Dudek et al. (1998) J. Comp. Chem. 19:548-573) indicated at 106A or ab initio secondary structure generation techniques of block 106B, techniques in which the alignments are adjusted using information on the secondary structure, functional residues, and disulfide bonds as described herein, can be used to complete the model (e.g. U.S. Patents Nos. 5,331,573; 5,579,250; and 5,612,895). This model, complete with loops, is then subjected to refinement procedures (block 110) based on molecular mechanics, molecular dynamics, and simulated annealing methods. Energetic refinement of the structure can be accomplished by performing molecular mechanics calculations using, for example, an ECEPP type forcefield (Dudek et al. (1998) J. Comp. Chem. 19:548-573) or through molecular dynamics simulations using, for example, a modified AMBER type forcefield (Ramnarayan et al. (1990) J. Chem. Phys. 92:7057-7076. As known to those of skill in the art a modified AMBER (version 3.3) force field is a fully vectorized version of

-32-

AMBER (3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy (see, e.g., Weiner et al. (1986) J. Comp. Chem. 7:230-252). If necessary, the 3-D structures can be dynamically refined, for example, by using a simulated annealing protocol (e.g.,, 100 ps equilibration, 500 ps dynamics, up to 1000°K, 1 fs data collection).

The refinement process step 110 is used to offset problems that may arise when homology models are not built carefully or when they are built using fully automated methods. Problems that may arise include chain breaks (e.g. consecutive C^a atoms are farther apart than the optimum distance of 3.7 to 3.9 Å); distorted geometry (e.g. bond lengths and bond angles are too far from their optimal values); cis-peptide bonds (e.g., incorrect isomerization of the peptide backbone in non-proline residues when it is not required); disallowed backbone and side-chain conformations (e.g., dihedral angles do not satisfy the Ramachandran plot (see, Balasubramanian (1974) Nature 266:856-857) criteria for a fully favorable protein structure conformation); and misfolded loops (e.g. nonhomologous loops are generated in unnatural conformations). The refinement procedure 110 removes distortions of covalent geometry by using energetic methdods, converts disallowed backbone and side-chain conformations into allowed ones using simulated annealing methods, conserves protein core structure and secondary structural elements built by homology, and rebuilds unnatural loop constructions (Dudek et al. (1998) J. Comp. Chem. 19:548-573).

15

20

For quality control (block 112), the protein structural characteristics, for example, stereochemistry (e.g.,, phi/psi and side chain angles), energetics (e.g.,, strain energy), packing profile (e.g.,, packing factor per residue) and hydrophobic packing are evaluated and required to meet acceptable criteria before the structures are used in further studies

_ 5

10

20

25

or inputted into a structural polymorphism database. Quality control using strain energies entails computing normalized residue energies (NREs) based on the equation:

$$e_i = [E(i,X) - E_{AV}(X)] / E_{SD}(X)$$
, where

E(i,X) is the energy of interactions of amino acid X in position i with protein environment and solvent;

 $E_{\rm AV}(X)$, $E_{\rm SD}(X)$ is the average residue energies and their standard deviations calculated for 20 amino acids in more than 100 high-quality crystal structures; and

NREs characterize how favorable the interactions of each residue are within the protein environment (Majorov and Abagyan, (1998) Folding & Design 3:259).

The average NRE characterizes the overall quality of a protein structure and is defined as:

15 $e_{av} = (1/N) \Sigma_i e_i$, where

 $e_{av} \le 0.5$ denotes high-resolution X-ray crystal structures;

 $e_{av} \leq 1.0$ denotes good as NMR and theoretical models; and

 $e_{av} \ge 1.5$ denotes structures that require further refinement.

After the quality of structure is determined at block 112, the model is checked at block 114 to determine if it is satisfactory. If the overall quality of the model is not satisfactory, a "No" outcome at block 116, then remedial action is undertaken to fix problems at block 118, including further iterative energy refinement (block 110), and repeated checking (block 114). The refinement and evaluation is repeated until the model is considered to be acceptable, a "Yes" outcome at block 120, whereupon structural and/or physical properties (e.g. energetics and phi/psi angles) are calculated at block 122A and clinical data (if available) is obtained at block 122B. The model is then inputted into a structural polymorphism database at block 124.

FIG. 2 shows an exemplary method for generating structural variant models derived from genetic polymorphisms and using them in structure-based drug design studies. At the block numbered 200, patient data is acquired for a gene that exhibits genetic polymorphisms. Protein sequence information is then derived, at block 202. A check is made for determination of the 3-D structure of the native protein. If the 3-D structure has been determined, a "Yes" outcome at block 206, then a multiple sequence analysis is performed at block 208 to determine the exact amino acid variations for the structure. If the 3-D structure has not been determined, a "No" outcome at block 210, then the structure is determined using physiochemical methods at block 212.

10

15

20

25

Next, at block 214, the 3-D structural models for all variants are generated. A refinement process is then completed at block 216 for the structural models. As noted above in connection with FIG. 1, the process involves subjecting each model, complete with loops, to refinement procedures based on molecular mechanics, molecular dynamics, and simulated annealing methods. As before, the energetic refinement of the structure can be accomplished by performing molecular mechanics calculations using an ECEPP type forcefield (Dudek et al. (1998) J. Comp. Chem. 19:548-573), or through molecular dynamics simulations using, for example, a modified AMBER type forcefield (Ramnarayan et al. (1990) J. Chem. Phys. 92:7057-7076), where a modified AMBER (version 3.3) force field is a fully vectorized version of AMBER (3.0) with coordinate coupling, intra/inter decomposition, and the option to include the polarization energy as part of the total energy (Weiner et al. (1986), J. Comp. Chem. 7:230-252). If necessary, the 3-D structures can be dynamically refined, for example, by using a simulated annealing protocol (e.g.,, 100 ps equilibration, 500 ps dynamics, up to 1000°K, 1 fs data collection).

-35-

At block 218, a quality evaluation is performed for all the models. As described in connection with the quality evaluation process in Fig. 1, the evaluation at block 218 involves evaluating the protein structural characteristics, for example, stereochemistry (e.g., phi/psi and side chain angles), energetics (e.g., strain energy), packing profile (e.g., packing factor per residue) and hydrophobic packing, which must meet acceptable criteria before the structures are used in further studies or inputted into a structural polymorphism database.

After the model quality is determined, at block 220 the models are checked to determine if they are satisfactory for further use. If a model is not satisfactory, a "No" outcome at block 222, then the problems are identified and solved with remedial action at block 224. The remedial action may include further iterative energy refinement at block 216 and repeated checks of model quality at block 218. Once the models are satisfactory, a "Yes" outcome at block 226, structure-based drug design methods are applied at block 228 to identify potential new drugs that bind to the structural variant models. The drug design methods are described further below.

10

15

20

25

FIG. 3 shows another exemplary and alternative method for generating structural variant models derived from genetic polymorphisms and using them in structure-based drug design studies. The process of FIG. 3 is similar to the process of FIG. 2 from the initial process at block 300 of acquiring patient data for a gene that exhibits genetic polymorphisms through the process of obtaining models that are satisfactory (a "Yes" outcome at block 326). Thus, block numbers in FIG. 3 from 300 through 326 that correspond to FIG. 2 blocks numbered from 200 thorough 226 refer to similar operations. Unlike FIG. 2, however, the process illustrated in FIG. 3 then involves docking operations.

-36-

At block 328, once the models are determined to be satisfactory, drug molecules are docked with the structural variant models. Next, at block 330, the free energy of binding is evaluated with the potential drugs under study for each structural variant model. At block 332, the total free energy of binding is decomposed, based on the interacting residue in the protein active site. Lastly, at block 334, the free energy of binding is correlated with patient data, if the data is available. Thus, the 3-D structural data is employed in drug design. Details of using such structural data in drug design are described further below.

c. Crystal structures

10

15

20

25

The crystal structure of any protein can be determined empirically and the resulting coordinates used as the basis for determing structures of variants. Such structures are often known (see, e.g., Kohlstaedt et al. (1992) Science 256:1773-1790 for a crystal structure of HIV-1 RT bound to a ligand).

3. Use of 3-D structural variant models in drug design

The structural differences in protein structural variants that arise due to genetic polymorphisms can have profound effects on biological activity. Because of the structural differences among the variants, they may have different physical or reactive properties and therefore may exhibit different biological activities. These differences may include, for example, different responses to a given drug, so that a drug which works well in a patient with one particular genetic polymorphism may not work as well in another patient exhibiting a different polymorphism.

The 3-D molecular structures of drug targets derived from genetic polymorphisms can be used in structure-based drug design studies to greatly advance the development of new pharmaceuticals. Relational databases of these 3-D structures that are derived from samplings of genetic polymorphisms over a patient population or a cross-section of the

-37-

population can be used to design potential drugs in order to optimize effectiveness for the particular population.

The structures and databases described herein can provide information that is useful, for example, in designing a drug that is effective in the greatest percentage of the population. It is desirable that a given drug is effective in the largest percentage of the population, since such a drug is likely to have the greatest clinical utility and thus the greatest commercial value. A drug with superior performance properties is sometimes referred to as a "best in class" drug and is highly prized by pharmaceutical companies since this heralds market leadership and the likelihood of commercial success. The databases and methods described herein can be used to determine 3-D protein structures for drug targets that are associated with particular genetic polymorphisms and to use the structures in drug design studies for design and optimization of candidate drugs that exhibit activity over the broadest patient population.

10

15

20

25

Genetic polymorphisms may result in target protein structural variants in which drug efficacy correlates with specific populations or subpopulations. In some cases, it might be desirable to target drug design or drug therapy toward a specific patient population, such as a particular race, gender, or age group, affected by a certain disease or condition or toward those having a specific genetic polymorphism. The information derived from comparing the 3-D structural variants arising from different genetic polymorphisms may be useful for understanding why drugs are active or inactive in different subpopulations, or for assisting in developing new drugs to maximize efficacy across specific populations.

-38-

a. Selection of relevant structural variants

The structural variant models in the structural polymorphism database provided herein can be used to design new drugs or to select a drug therapy that would be appropriate for a patient exhibiting a particular genetic polymorphism. As it may not be possible for a drug to work equally well for all polymorphisms, and thus all patients, representative structural variants can be selected for use in drug design studies in order to maximize biological activity based on genetic polymorphisms.

10

15

20

25

In some cases, structural variants are analyzed to determine the common structural features that are conserved through the selected models. These conserved features are used as a basis for drug design. In some cases, the structural variant corresponding to the genetic polymorphism occurring most commonly in a population can be selected for use in identifying drugs that would be effective in the greatest percentage of the population. Optionally, structural variants corresponding to a relevant subpopulation, such as a particular gender, age, race, or other characteristic, can be selected for use in designing drugs that are active in that subpopulation. In other cases, individual structural variant models can be selected for use in designing drugs that are specifically active against one target in one individual arising from a particular genetic polymorphism. Additionally, model structures that represent variants derived from patients that receive a specific treatment regimen or exhibit a particular clinical response (e.g. drug resistance) to a given drug are used as bases for drug design.

The relevant structural variants may be identified using the structural analysis tools described herein, optionally in combination with database and statistical analysis tools that permit a complete analysis and comparison of the molecular structures and properties of the structural

-39-

variants. The structural variants selected based on the criteria including, but not limited to, those listed above are used in drug design.

b. Drug design

10

15

20

25

Once the protein target structural models have been selected, structure-based drug discovery methodologies, for example, computational screening or docking (e.g., DOCK (available from University of Ca, San Francisco; and AUTODOCK available from Scripps Research Institute, La Jolla and others referenced herein or known to those of skill in the art), can then be used to design biologically-active compounds based on the 3-D structures of the biomolecular receptors.

Using these methods, drug designers can identify and computationally rank various potential clinical drug candidates for maximum efficacy, thus cutting the time and expense associated with drug discovery. The preferred design of drug candidates or the modification of existing drugs is based on the intermolecular interactions between the drug candidate or modified drugs and the selected structural variants predicted by computationally docking drug molecules with the target protein models; energetically refining the docked complexes; determining the binding interactions between the drug or potential new drug candidate molecules and the models by calculating the free energy of binding of the docked complexes and decomposing the total free energy of binding based on interacting residues in the protein active site or sites deemed important for protein activity.

c. Computational docking

Methods for using the structural variant models to design potential new drugs or to aid in the selection of a drug therapy based on the interactions of selected small molecules with the particular variants are provided. Structure-based drug design experiments, such as computational screening or docking studies, calculation of binding

-40-

energies or analysis of steric, electrostatic or hydrophobic properties of the resulting structural variant models, can be performed on selected structural variant models to aid in the understanding of observed biological activities or to determine new potential drug candidates to bind to the particular target.

In a typical computational docking protocol, the active site, or sites deemed important for protein activity, of the protein model is defined. A molecular database, such as the Available Chemicals Directory (ACD) or any database of molecules, is screened for molecules that complement the protein model. Solvation parameters are factored in (see, e.g., Shoichet et al. (1999) PROTEINS: Structure, Function, and Genetics 34:4-16). In these computational docking studies, drugs or drug candidates are fitted to the structural variant models based on complementary interactions (e.g., steric, hydrophobic, or electrostatic interactions).

Methods for performing such studies are well known and software tools for performing the calculations are widely available (M. Lambert, "Docking Conformationally Flexible Molecules into Protein Binding Sites" in Practical Application of Computer-Aided Drug Design, Charifson, Ed., Marcel Dekker, NY, pp. 243-303; Kurtz (1992) Science 257:1078-1082; Kuntz

et al. (1982) J. Mol. Biol. 161:269-288; Stewart et al. (1992) Med. Chem. Res. 1:439-443; Shoichet et al. (1993) Science 259:1445-1450; Shoichet et al. (1991) J. Mol. Biol. 221:327-346).

New potential drug candidates can be designed by identifying potential small molecule drugs that can bind to a particular structural variant. This is accomplished, for example, by methods including, but are not limited to, methods for electronic screening of small molecule databases as described herein, methods involving modifying the functional groups of existing drugs in silico, methods of de novo ligand design. Methods for computationally designing drugs are known to those

25

of skill in the art and include, but are not limited to, DOCK (Kuntz et al. (1982) "A Geometric Approach to Macromolecule-Ligand Interactions", J. Mol. Biol., 161:269-288; available from University of Ca, San Francisco); and AUTODOCK (see, Goodsell et al. (1990) "Automated Docking of Substrates to Proteins by Simulated Annealing", Proteins: Structure, Function, and Genetics, 8, pp. 195-202; available from Scripps Research Institute, La Jolla); GRID (Oxford University, Oxford, UK); CAVEAT (UC Berkeley, Ca), LEGEND (Molecular Simulations, Inc., San Diego, CA); LUDI (Molecular Simulations, Inc., San Diego, CA); HOOK (Molecular Simulations, Inc., San Diego, CA); CLIX (CSIRO, Australia); GROW 10 (Upjohn Laboratories, Kalamazoo); others including HINT, LUDI, NEWLEAD, HOOK, PRO-LIGAND and CONCERTS (see, M. Murcko, "An Introduction to De Novo Ligand Design" in Practical Application of Computer-Aided Drug Design, Charifson, Ed., Marcel Dekker, NY, pp 305-354), methods based on QSAR (quantitative structure-activity relationships, QSAR and Drug Design: New Developments and Applications, Fugita, Ed., (1995) Elsevier, pp 3-81; 3D QSAR in Drug Design, Kubinyi, Ed., (1993) Escom, Leiden), and other methods known to those of skill in the art for determining molecules that have optimal binding interactions with a selected target. 20

The docked complexes, if needed, are further refined energetically to optimize geometries within the binding site and to select the best structure from a set of possible structures, using molecular mechanics, molecular dynamics, and simulated annealing techniques, including those described herein and others that are known to those skilled in the art.

d. Free energy of binding studies

After the computational docking step, the free energy of binding of the docked complex is calculated, and the total free enegy of binding is decomposed based on the interacting residues in the protein active site or

sites deemed improtant for protein activity. Analyses of the binding energies are needed to identity drug candidates. If need or desired, the free energy of binding of different drugs or potential drugs to each structural variant model can be calculated by substracting the free energy of the non-interacting protein and drug from the free energy of the protein-drug complex. The total free energy of binding is decomposed into its various thermodynamic components, e.g. enthalpic and entropic components, based on the interacting residues in the protein active site in a solvated model to characterize the structural and thermodynamic 10 features in the mode of drug binding and to determine the contribution of the solvent] (see, e.g., Wang et al. (1996) J. Am. Chem. Soc. 118:995-1001; Wang et al. (1995) J. Mol. Biol. 253:473-492; Ortiz et al. (1995) J. Med. Chem. 38:2681-2691, which describes a computational method for deducing QSARs from ligand-macromolecule complexes). Following the computational drug design protocol described herein, any potential new drugs that are identified can be synthesized in, for example, industry or academia, and subjected to further biological testing, such as in vitro studies or pre-clinical and clinical in vivo testing.

Based on the predicted intermolecular interactions of the drugs or modified drugs with the structural variant models from binding studies, potential drug candidates that are specific for a protein with a selected polymorphism or that specifically interact with all proteins exhibiting the polymorphism can be identified.

15

20

25

It is also possible to individualize drug design or drug therapy by determining the structural variants associated with a particular patient and then designing or screening drugs or potential drugs to maximize efficacy in that subject or in a subpopulation that exhibits the same genetic polymorphism. The variants may also be used to track polymorphic variations in infectious organisms, such as viruses. For example, the

human immunodeficiency viruses (HIVs) reverse transcriptase and protease have served as drug targets (see, Erickson et al. (1996) Ann. Rev. Pharmacol. Toxicol 36:545-571); their three-dimensional structures are known (see, e.g., Nanni et al. (1993) Perspectives in Drug Discovery and Design 1:129-150; Kroeger et al. (1997) Protein Eng. 10:1379-1383). The clinical emergence of drug-resistant variants of these viruses has limited the long-term effectiveness of drugs targeted against these enzymes.

10

20

25

As noted, these enzymatic proteins in order to preserve function must exhibit conserved 3-D structures. The methods herein permit design of drugs specific for the conserved regions of the 3-D structures. They also permit selection of drug regimens based upon the alleles expressed. Hence, methods for designing HIV enzyme-specific drugs are provided. Flow charts illustrating exemplary alternative embodiments using protein 3-D structures derived from genetic polymorphisms in structure-based drug design studies are provided (see, Figs. 2 and 3). In the flow charts depicted in these figures, the drug design includes structure-based drug design methods (see, Figure 2) and computational docking of drugs with structural variants, evaluation of the binding energy of the docked complexes, and correlation of the binding energy with patient data such as age, gender, race, drug treatment history, and any other pertinent information that is available (see, Figure 3). The data generated by this computer-based method can be stored in a database, such as, for example, in a relational database. The resulting database can be screened using searching tools to select potential drugs and therapeutic agents that bind to or exhibit biological responses towards target proteins.

C. Applications of computer-based methods

... **5**

20

25

As discussed above, the computer-based methods provided herein include some or all of the steps of obtaining one or more, preferably two or more, amino acid sequences of a target protein that is the product of a gene exhibiting genetic polymorphisms; generating 3-dimensional (3-D) protein structural variant models from the sequences; and based upon the structures of the 3-D models, designing drug candidates or modifying existing drugs based on the predicted intermolecular interactions of the drug candidates or modified drugs with the structural variants by 10 computationally docking drug molecules with the target protein models; energetically refining the docked complexes; determining the binding interactions between the drug or potential new drug candidate molecules and the models by calculating the free energy of binding of the docked complexes and decomposing the total free energy of binding based on interacting residues in the protein active site or sites deemed important for protein activity. There are numerous applications of these methods, which include structure-based drug design and drug testing; selection of clinically relevant populations for drug testing and other such methods.

Genetic polymorphisms and structure-based drug design

As noted above, structure-based drug design is an increasingly useful methodology that has made a great impact in the design of biologically active lead compounds. Drug designers can design and screen potential new drugs via computational methods, such as docking or binding studies, before actually beginning patient testing. The drugs designed by such methods, and also those identified by traditional methods of drug discovery, are then tested in clinical trials. Among those that show efficacy for a particular indication and low toxicity ultimately are approved for use. It is found, however, that not all patients with a

PCT/US00/30863 WO 01/35316

-45-

particular indication respond uniformly to the drugs. The drug may not be efficacious or side-effects may be pronounced.

The methods provided herein, represent a further advance in the use of rational drug design methods. As described herein, polymorphic variation has an effect upon the 3-D structure of encoded proteins. As a result, drugs interact with variants differently, leading to differential responses in the population as a whole. A new approach to drug design and testing is provided herein. This methods involves identifying polymorphisms and determining 3-D resulting structures, which are then used in methods, including, computational drug design, in the selection of patient populations, in designing treatment protocols and in other applications.

2. Drug resistance

5

10

20

Methods for understanding and overcoming drug resistances by using 3-D protein model structures resulting from multiple genetic polymorphisms or mutations in an infectious agents, such as viruses, bacterial and other pathogenic agents are provided. Also provided are methods that for using this information in drug design studies.

In the case of infectious organisms or other replicating or mutating agents, such as flu, HIV, rhinovirus or biological warfare agents, some polymorphisms or mutations may arise over time which convey resistance or susceptibility to specific drug therapy, for example, by altering the drug target structure or physical properties so that a specific drug or therapy, such as an antibiotic or vaccine, may no longer be able to bind to or 25 otherwise interact with the target protein to exert its desired biological effect. For certain infectious agents, such as HIV, genetic polymorphisms in certain genes give rise to drug resistance as the virus mutates (see, e.g., Erickson et al. (1996) Annu Rev. Pharmacol. Toxicol. 36:545-571).

-46-

Where drug resistance that arises from mutations or polymorphisms is observed, the methods described herein can be used to develop new drugs that overcome the resistance. For example, once drug resistance is observed, the structure associated with the resistant polymorphism can be determined and used in further drug design studies to suggest new drugs or modifications to the existing drug that will restore biological activity by targeting different mutants or that will target multiple mutants simultaneously.

The model structures can also be used to correlate drug resistance in infectious diseases with the structural variants derived from genetic polymorphisms. Here, the 3-D structure of the virus or other drug target is determined for the particular variant model against which the drug was effective. When drug resistance arises due to a genetic polymorphism, a model for the structure variant associated with the resistant organism can be generated, and a new drug can be designed or modifications can be made to the existing drug to overcome the resistance.

10

15

20

For example, samples of the mutating organism can be obtained over time and structural models for the resulting proteins can be generated. These models can then be used to design new drug therapies that are active against the mutated organism. Multiple drug resistant structures can be analyzed to obtain an average structure or to identify common structural features in order to design new drugs that have the broadest spectrum of activity against multiple mutations.

Such structural information is useful in designing effective drug
therapies to overcome resistance or to develop drugs that are effective
over a range of genetic polymorphisms and thus work for the maximum
number of patients.

-47-

3. Identification of conserved structural features or pharmacophores

If common structural features are observed over a range of protein targets that are derived from genetic polymorphisms, these common features may be used to design a drug that is effective with a variety of genetic polymorphisms and thus many patients. The retention of certain common structural features over a large number of genetic polymorphisms suggests that those features may not be mutatable because the conserved structure may be essential to protein function,

10 e.g., to the viability of an infectious organism or virus. Such conserved structural elements are prime targets for structure-based drug design, e.g., anti-infective or antibiotic drug design, and can lead to highly effective therapies.

The common structural features can serve as a basis for structure-based drug design, for example, by serving as a scaffold for building a receptor model into which potential drug candidates can be docked or as a pharmacophore query for screening a library of physical or virtual chemical or biochemical molecules to identify compounds that match the pharmacophore template and, thus, are potential drug candidates.

15

20

25

30

Analysis of 3-D protein structural variants derived from genetic polymorphisms to identify the common structural features over a large number of structural variants can aid in the design of drugs that are active over a broad range of genetic polymorphisms, such as in a large number of patients or against drug resistant targets.

In comparing sets of related protein structures, such as those with the same biological function or those resulting from genetic polymorphisms, certain parts of the structural framework are often found to be conserved, while other parts vary among the proteins. Mutations that occur in the conserved regions of the structure can have significant effects biological activity. For example, in viruses, the conserved features

-48-

can be essential to protein function and, thus, to the viability of the infectious organism or virus. Identifying the conserved structural features over a range of structures often gives insight into which structural features are necessary for biological activity and are therefore non-mutatable. By analyzing a number of structural variants derived from genetic polymorphisms that exhibit drug resistance, it is possible to identify or design drugs that interact best with the common structural features in all of the variants. Using these features in structure-based drug design studies leads to the identification of drugs that retain biological activity despite multiple mutations, or polymorphisms, and could help to overcome the problem of drug resistance.

In certain preferred embodiments, new potential drug candidates can be identified using the structural variant models by identifying pharmacophores or conserved features in the protein structural variant models and using this structural information to identify small molecules that would bind to the structural variant models.

10

15

20

25

Using structural comparison tools described herein, the common structural features that are conserved across a range of structural variant models of a given protein based on different genetic polymorphisms can be identified. To do this, multiple structural variant models are compared, generally by superimposing the coordinates of one variant model onto those of one or more other variants and observing the structural fit. Such functionality is commonly found in molecular graphics or homology modeling packages. Once the optimum fit of structures is performed, then the structural features that are present throughout the structural variant models can be identified and used as the basis for drug interactions in structure-based drug design studies. For example, the pharmacophores or conserved features can be specified as database queries and a library or database of small molecule structures can be

___5

10

15

20

25

searched to identify new lead compounds to bind to the pharmacophores. Alternatively, other structure-based ligand design strategies can be employed to design lead compounds or to identify modifications to be made to existing drugs to improve biological activity.

4. Identification of compensatory structural changes

Certain proteins, for example, viral proteins or other infectious organisms, may harbor multiple genetic polymorphisms. Since each genetic polymorphism can give rise to slight changes in structure, some, and over time, many, additional genetic polymorphisms may cause changes in the protein structures that significantly affect biological activity. These structural changes could result in, for example, different dynamical behavior, alteration in enzyme kinetics or differences in substrate recognition, which can significantly alter drug response. For example, a mutation for one drug compound can suppress a mutation to a second drug due to compensatory effects. In these cases, a drug which is predicted to be ineffective for a given patient based upon the single nucleotide correlation may, in fact, be effective as a result of these changes.

Because mutations are so frequent in AIDS and other viruses, few sequences are exactly the same in different patients. Thus, it is difficult or inconclusive to generate multiple mutation sequence correlations for drug resistance. If each patient has a different viral sequence due to a high viral mutation rate, then no sequence correlation is even possible in such cases.

The methods described herein can be used to study the effects of multiple genetic polymorphisms on a resultant protein structure. Multiple mutations are common in AIDS and other viruses, which makes sequence correlation difficult. By observing the structural effects of the mutations on the resulting protein, it is possible to look at the net effect of all

structural changes and to consider the overall structure of the protein in drug design studies. For example, a mutation might occur in the active site, or site of drug action, in a protein. Additionally, there may be related mutations in other parts of the protein structure, which might not be identified from a single point mutation correlation. These related mutations could have an effect on biological activity of the protein. By looking only at the active site, it might be predicted that a drug or potential drug would not bind to the protein. The additional mutation, however, might cause compensatory structural changes in the protein structure that alter its properties in a way that restores biological activity.

By computing 3-D protein structures from gene sequences containing multiple polymorphisms, it is possible to more accurately predict the effect of multiple sequence mutations on protein structure and, thus, to obtain a better correlation between sequence and drug resistance than by considering sequence correlations alone. This information can be useful, for example, in understanding drug resistance and can aid researchers and clinicians in developing new drug therapies to overcome drug resistance.

The structures that are derived based on multiple generic polymorphisms can be used in structure-based drug design studies to provide frameworks, or scaffolds, into which drug or potential drug molecules can be docked. This permits the design of drugs that are active against a wider range of structural variants, thus, in more patients or against a range of drug resistant proteins.

5. Clinical Applications

5

10

15

20

25

A knowledge of the repertoire of structural differences arising from genetic polymorphisms across the human population or specific subpopulations can provide insight into the differing biological responses in patients based on their genetic differences. For example, where clinical

-51-

data are available for patients having particular genetic polymorphisms, this information can be associated with the 3-D protein structural variants and used to find correlations between polymorphisms and observed drug responses.

The methods provided herein can be used to design drug therapies that bring about favorable clinical responses (or eliminate unfavorable effects) in patients, to identify pharmacological effects of drugs in different patient subpopulations (e.g. age, race, gender) and to simulate clinical trails to increase the probability that the trials will yield optimal results.

5

10

15

20

Because of the high cost of clinical trials, such studies are generally focused on small patient populations. The structural analysis tools described herein permit the extension of clinical trials to cover patient populations not specifically included in the study. This is accomplished through correlation of the structural variants derived from genetic polymorphisms with clinical responses.

The molecular structures and databases described herein can also find application in the understanding and prediction of clinical or pharmacological drug responses, for example, efficacy, toxicity, dose dependencies or side effects in patients. For example, relational databases containing 3-D protein structural variants can provide a means for managing and using the information to understand and predict clinical responses in patients.

In other embodiments, observed clinical data from patients in a clinical trial can be associated with the structural variant models for each genetic polymorphism exhibited in the clinical subjects, for example, in a structural polymorphism relational database. The correlation between the structural variants and observed clinical effects can then be utilized to predict clinical outcomes in patients that did not participate in the clinical

-52-

trial. For example, a structural variant model can be generated for a patient based on a genetic polymorphism exhibited in the patient, and the database can be mined to identify structurally similar variants for which clinical results are known. Structural similarity can be determined, for example, by superimposing the structures and measuring the RMS (root mean squared) differences between the structures or by using pattern matching or motif searching algorithms. The results can be used to predict clinical responses in the patient based on the clinical data associated with the structurally similar variants.

The predicted correlations can also be used to aid in the design of subsequent clinical trials. The follow-on trials can be made more effective through the judicious selection of patients with given genotypes (i.e., those exhibiting the same genetic polymorphisms), as guided by the structurally predicted outcomes. For example, a clinical trial can be designed based on a subpopulation of clinical subjects which exhibit a specific genetic polymorphism (i.e. structural variant) to demonstrate the effectiveness of a given therapeutic on a targeted population.

10

15

20

25

In other embodiments, the methods provided herein can be used in the selection of drug therapies for patients exhibiting a particular genetic polymorphism. This is accomplished by generating the structural variant model associated with the polymorphism, docking drug molecules that might be used to treat the patient into the structural variant model and calculating the binding energies of each drug with the variant. The results of docking or free energy calculations can be correlated to clinical data, for example, patient population (e.g., ethnic background, race, sex, age), treatment regimen, patient response to a particular drug or duration of treatment. The binding energies can be compared, for example, to determine which drug would best bind to the variant in order to identify

the drug that could best be used to treat the patient to optimize biological activity.

D. Creation of 3-D Structural Polymorphism Databases

5

10

15

25

The above-noted methods all rely upon the use of databases of nucleic acid sequences. Any such database known to those of skill in the art may be employed; numerous such databases are publically available (e.g. the Stanford HIV database). The Stanford HIV database is hierarchal database with information about HIV patients who received or did not receive protease inhibitor treatments, patient-dates, isolates, sequences, hyperlinks to MEDLINE and GenBank abstracts, and art. This database, however, does not contain 3-D protein structures of any proteins including HIV reverse transcriptase (RT) and HIV protease (PR; see, e.g., Shafer et al. (1999) Nucleic Acids Res. 27:348-352, Shafer et al. (1999) J. Virol 73:6197-6202, http://hivdb.stanford.edu/hiv, Richter (January 20, 1999) "AIDS drugs found to be effective in the world's most common HIV strains).

Databases of sequences and associated information may also be generated as described herein by obtaining samples and sequences from a variety of sources. In all instances, further databases are generated by then calulating 3-D structural models of the encoded proteins or relevant portions, such as active binding sites, thereof, from the nucleic acid sequence information. It is these databases of nucleic acid sequence and/or primary protein sequence and the associated 3-D structure that are provided herein and that are used in the all of the methods, except for the computational phenotyping discussed below, which does not require a database, provided herein. Hence databases comtaining computationally determined 3-D structures of polymorphic proteins or portions thereof are provided herein. These databases serve as tools in a variety of methods, including those provided herein.

-54-

Databases that include 3-D structures for variant proteins encoded by the nucleic acids that contain polymorphisms are provided. These are generated after 3-D structural models are constructed for the protein structural variants, preferably for all of the protein structural variants, representing the genetic polymorphisms, by inputting the atomic coordinates into a structural polymorphism database, preferably a relational database, and optionally with associated structural and/or physical properties (e.g., phi/psi and side-chain angles and energetics), and other data, if available, including, but are not limited to, historical 10 data, such as parental medical histories, and clinical data. The resulting database is used in structure-based drug design studies and for clinical analyses. Figure 11 is a tabulation of the 3-D coordinates of a representative entry, an HIV protease, that is encoded by the DNA in one of SEQ ID Nos. 3-74 and 77-117, and that is an entry in an exemplary database that includes 3-D structures. Exemplary databases that contain the nucleic acids sequences and structures of all proteins encoded by SEQ ID Nos. 3-117 as well additional nucleic acids are provided herein and are described in the EXAMPLES.

A database is preferably interfaced to a molecular graphics package
that includes 3-D visualization and structural analysis tools, to analyze
similarities and variations in the protein structural variant models (see,
copending U.S. application Serial No. 09/531,995, which is published as
International PCT application No. WO 00/57309, and is a continuation-inpart of U.S. application Serial No. 09/272,814, filed March 19, 1999).

Briefly, International PCT application No. WO 00/57309 provides a
database and interface for access to 3-D molecular structures and
associated properties, which can be used to facilitate the design of
potential new therapeutics. The interface also provides access to other
structure-based drug discovery tools and to other databases, such as

databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites. The interface also provides access to other structure-based drug discovery tools and to other databases, such as databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites. This interface can be modified as needed to adapt for use with a paritcular database.

10

15

20

25

A relational database that collects multiple data files relating to the same molecular structure in the same subdirectory and that provides an interface to access all of the collected files from the same structure using the same user interface program is also provided. The collected files include a variety of information and computer file formats, depending on the type of information to be conveyed to users of the database. In practice, a user communicates over a public network, such as the Internet, or over a controlled network, such as an internet, with a secure file server that controls access to the collected files, and the interface to the collected files is provided by a standard graphical user interface program that is widely available. In this way, a convenient means of searching molecular structure data for characteristics of interest is provided. Data searching, file viewing, and investigation of multiple representations of molecular structures from within a single viewing program can also be performed using the database and interface.

The data files can be those available over a wide network such as the Internet, and a suitable graphical user interface designed or obtained. Such interface is used for viewing the data files is a standard Internet web browser program, such as the web browser products by Netscape Communications, Inc. and Microsoft Corporation that are distributed free

-56-

of charge. Such browser products readily import and provide views of files having a wide variety of formats that contain alphanumeric, video, and audio data. A security server is preferably located between the user browser program at a network client machine controls access to the database, which is housed at a file server connected to the security server. Before a user gains access to the database, the security server checks authorization for the individual user and then, if appropriate, permits downloading of appropriate data from the database file server. It is contemplated that the databases containing 3-D structures of proteins or portions thereof the exhibit polymorphism will be loaded.

10

15

20

Data for a molecular structure is loaded into the database by specifying the file pathnames for the various data files that contain the different types of data, including the different molecule views. Using a browser to view the data files permits various helper applications, called plug-ins, to smoothly and transparently accept the different file formats and provide views to the user. The various data files of the database are organized in accordance with the database design when they are loaded into the database and are managed by a relational database management program.

In addition to 3-D protein structures and associate primary sequences, as provided herein, the database can optionally contain associated biological or clinical data, such as drug resistance, side effects, efficacy, pharmacokinetics and other data, that correlate with or can be correlated the structural variants. This information will be used for correlating observed clinical effects to specific structural variants and for predicting clinical responses and outcomes based on a patient's structural variants, *i.e.*, genetic polymorphisms.

10

15

20

Structural analysis tools are preferably integrated with the structural database for comparing and analyzing the resulting protein structural variant models. For example, the molecular graphics software package described in International PCT application No. WO 00/57309, includes structural analysis capability to measure the structural attributes of the model (distances, angles, etc.), to analyze sequences and secondary structures, to study physical properties such as hydrophobicity, electrostatic potential, and active or reactive sites in the protein, as well as to evaluate the quality of the structure (both conformationally and energetically).

Structures can also be compared by aligning them, such as by performing a least squares fitting of the x-, y- and z-coordinates of each of the structural variant models and superimposing the structures or any other alignment method or structural comparison method. For example, the structures of the variants can be clustered, or grouped together, based on structural similarity. This can save time over studying each structural variant independently because, where structures are considered to be similar enough that they are clustered together (e.g., if their structures can be superimposed within a specified tolerance), then only a representative structure, or perhaps an average structure or scaffold, which is derived as a composite of the individual structural variant models, can be used in further drug design studies.

Tools for database searching can also be included in the software package. These can be used to query the database for structural variant models having similar properties, such as molecular structure or sequence similarity. These tools are used, for example, to mine the database to identify variant models that are structurally similar (e.g. to find structures that overlap within a specified tolerance), and thus would be predicted to interact in the same way with potential drugs or exhibit the same clinical

-58-

response. This information could be useful in understanding the structural or clinical effects of different genetic polymorphisms and could potentially save time and money by extending the results of previously performed clinical or computer-based drug design studies to predict the results of studies on similar structural variants that have not yet been performed.

1. Exemplary Databases

. 5

10

15

20

25

Databases containing data representative of the 3-D structure of structural variants encoded by a selected gene or genes or the 3-D structure of other polymorphic variants are provided. The selected genes can be drug target, such as receptors and genes of infectious agents, such as the HIV protease or reverse transcriptase. Exemplary databases are presented in Example 5 which describes the construction, interface, use and appliations of HIV PR and RT databases. These databases may be stored on any suitable medium and used in any suitable computer system. Systems and methods for generating, storing and processing databases are well known.

2. Computer systems

Computer systems for processing the databases and computer systems containing the databases are provided. The processing that maintains the database and performs the methods and procedures using the databases may be performed on multiple computers, or may be performed by a single, integrated computer. For example, the computer through which data is added to the database may be separate from the computer through which the database is sorted or analyzed, or may be integrated with it. Each computer operates under control of a central processor unit (CPU), such as a "Pentium" microprocessor and associated integrated circuit chips, available from Intel Corporation of Santa Clara, California, USA. A computer user can input commands and data from a

10

15

25

keyboard and display mouse and can view inputs and computer output at a display. The display is typically a video monitor or flat panel display device. The computer also includes a direct access storage device (DASD), such as a fixed hard disk drive. The memory typically includes volatile semiconductor random access memory (RAM). Each computer preferably includes a program product reader that accepts a program product storage device from which the program product reader can read data (and to which it can optionally write data). The program product reader can include, for example, a disk drive, and the program product storage device can comprise removable storage media such as a magnetic floppy disk, an optical CD-ROM disc, a CD-R disc, a CD-RW disc, or a DVD data disc. If desired, computers can be connected so they can communicate with each other, and with other connected computers, over a network. Each computer can communicate with the other connected computers over the network through a network interface (see, e.g., Examples below) that permits communication over a connection between the network and the computer.

The computer operates under control of programming steps that are temporarily stored in the memory in accordance with conventional 20 computer construction. When the programming steps are executed by the CPU, the pertinent system components perform their respective functions. Thus, the programming steps implement the functionality of the system as described above. The programming steps can be received from the DASD, through the program product reader, or through the network connection. The storage drive can receive a program product, read programming steps recorded thereon, and transfer the programming steps into the memory for execution by the CPU. As noted above, the program product storage device can include any one of multiple removable media having recorded computer-readable instructions,

-60-

including magnetic floppy disks and CD-ROM storage discs. Other suitable program product storage devices can include magnetic tape and semiconductor memory chips. In this way, the processing steps necessary for operation can be embodied on a program product.

5

10

20

Alternatively, the program steps can be received into the operating memory over the network. In the network method, the computer receives data including program steps into the memory through the network interface after network communication has been established over the network connection by well known methods that will be understood by those skilled in the art without further explanation.

The computer that implements the client side processing, and the computer that implements the server side processing, or any other computer device of the system, may comprise any conventional computer suitable for implementing the functionality described herein. FIGURE 9 is a block diagram of an exemplary computer device 900 such as might comprise any of the computing devices in the system. Each computer operates under control of a central processor unit (CPU) 902, such as an application specific integrated circuit (ASIC) from a number of vendors, or a "Pentium"-class microprocessor and associated integrated circuit chips, available from Intel Corporation of Santa Clara, California, USA. Commands and data can be input from a user control panel, remote control device, or a keyboard and mouse combination 904 and inputs and output can be viewed at a display 906. The display is typically a video monitor or flat panel display device.

The computer device 900 may comprise a personal computer or, in the case of a client machine, the computer device may comprise a Web appliance or other suitable Web-enabled device for viewing Web pages. In the case of a personal computer, the device 900 preferably includes a direct access storage device (DASD) 908, such as a fixed hard disk drive (HDD).

The memory 910 typically comprises volatile semiconductor random access memory (RAM). If the computer device 900 is a personal computer, it preferably includes a program product reader 912 that accepts a program product storage device 914, from which the program product reader can read data (and to which it can optionally write data). The program product reader can comprise, for example, a disk drive, and the program product storage device can comprise removable storage media such as a floppy disk, an optical CD-ROM disc, a CD-R disc, a CD-RW disc, a DVD disk, or the like. Semiconductor memory devices for data storage and corresponding readers may also be used. The computer device 900 can communicate with the other connected computers over a network 916 (such as the Internet) through a network interface 918 that enables communication over a connection 920 between the network and the computer device.

10

15

20

25

The CPU 902 operates under control of programming steps that are temporarily stored in the memory 910 of the computer 900. When the programming steps are executed, the pertinent system component performs its functions. Thus, the programming steps implement the functionality of the system illustrated in FIGURE 1. The programming steps can be received from the DASD 908, through the program product 914, or through the network connection 920, or can be incorporated into an ASIC as part of the production process for the computer device. If the computer device includes a storage drive 912, then it can receive a program product, read programming steps recorded thereon, and transfer the programming steps into the memory 910 for execution by the CPU 902. As noted above, the program product storage device can comprise any one of multiple removable media having recorded computer-readable instructions, including magnetic floppy disks, CD-ROM, and DVD storage discs. Other suitable program product storage devices can include magnetic tape and semiconductor

-62-

memory chips. In this way, the processing steps necessary for operation in accord with the methods herein can be embodied on a program product.

Alternatively, the program steps can be received into the operating memory 910 over the network 916. In the network method, the computer receives data including program steps into the memory 910 through the network interface 918 after network communication has been established over the network connection 920 by well-known methods that will be understood by those skilled in the art without further explanation. The program steps are then executed by the CPU 902 to implement the processing of the system.

10

15

20

25

To implement the functionality described herein, it has been found that a suitable computer for performing database server tasks includes a "Pentium" level CPU having at least 128 MB of memory, 30 GB of disk storage, and 256 MB of disk swap space for files. A recommended configuration for computer performance would include, for example, a "Pentium III" processor at 700 MHz or faster, memory of 256 MB or greater, disk storage space of 50 GB or more, and swap space of 500 MB or more. A suitable configuration for performing user tasks as described above includes a "Pentium" level CPU having 128 MB memory, disk space of 240 MB with swap space of 256 MB, and an optional display circuit card supporting OpenGL and having 4 MB of memory. A recommended configuration includes, for example, a "Pentium III" processor at 500 MHz or faster, memory of 256 MB or greater, disk space of 500 MB or more, swap space of 500 MB or more, and an optional display card having 8 MB of memory or more, supporting resolution of 1024 x 768.

In a preferred embodiment, the software used in the computing system described above includes, for the server machine, operating system software such as "Windows NT Server 4.0" from Microsoft

25

Corporation, with Service Pack 5, Version 1280 (10 June 1999) or more recent, with database management server software such as, but are not limited to, "Oracle Server Standard Edition 8.1" from Oracle Corporation. The software used in a preferred embodiment of the user machine includes operating system software such as "Windows NT Workstation 4.0" from Microsoft Corporation, with Service Pack 5, version 1280 (10 June 1999) or more recent, as well as "Oracle Client Standard Edition Version 8.1" or higher. The client machine will also be compliant with the "Java" programming language (Java Runtime Environment 1.2.2). As will be known to those skilled in the art, other configurations may be suitable, depending on the applications being used and the computer performance desired.

E. Computational phenotyping

Also provided herein is a method designated computational phenotyping. Computational (also referred to herein as in silico 15 phenotyping). This refers to the method in which a 3-D protein structure is generated from a given genotype and protein-drug binding analyses in silico (computationally) are performed in order to determine whether drug binding does (i.e. sensitive) or does not (i.e. resistant) take place. This type of analysis is contemplated to be performed for an individual patient 20 or subject or groups thereof, such as ethnic groups, gender-based or agebased groups, particular species or groups thereof) to assess or select a drug for treatment of a particular disease or other such use, and is done to assess efficacy of a particular drug on a desired target, where the target exhibits polymorphisms. The following discussion and example, below, is with reference to HIV PR and RT, but it is understood that the methods and applications can be applied to any protein or gene product that exhibits polymorphic variation, and particularly to gene products that are drug targets.

-64-

Among the methods of computational phenotyping, there are three distinct methodologies that are clinically useful for determining either resistance or sensitivity to particular HIV-1 antiviral therapeutics. These are: genotyping, phenotyping, and *virtual* phenotyping. These methodologies are used to optimize the choice of therapeutics during the initiation of therapy, after drug failure, and/or during salvage therapy. Genotyping involves extracting the HIV viral RNA and amplifying all or part of the genes encoding the protease and reverse transcriptase proteins and sequencing them in order to assess the presence of resistance-associated mutations.

5

10

In phenotyping, the amplified sequences are instead sub-cloned into expression vectors and then tested for their replicative ability in vitro by transfecting them into cultured and/or established cell lines, such as, for example, human T cells, monocytes, macrophage, dendritic cells, 15 Langerhans cells, hematopoeitic stem cells, HeLa, XC, Mm5MT, LTL, COS 7, NIH3T3, LTA, MCF-7, or other cells derived from human tissues and cells that which are the principal targets of viral infection in the presence or absence of antiviral drugs (see, e.g., U.S. Patent No. 5,837,464; see, also EP 0852626; EP 1012334; and EP 0877937), 20 Virtual phenotyping (ViroLogic, Inc.) is an interpretive service in which the phenotype of a specimen (i.e. of a plant, animal, pathogen, or human) is inferred from the specimen's genotype based upon an extensive correlative database of known genotypes and phenotypes. Such a correlative database must be updated constantly to maintain clinical 25 accuracy.

Similar to *virtual* phenotyping, computational or in *silico* phenotyping infers phenotype based upon specimen genotype. Computational phenotyping is distinct from *virtual* phenotyping in that sensitivity or resistance to drugs is determined directly through protein-drug binding

analysis performed in silico and not through correlation with a database of known genotypes and phenotypes. The advantage of computational phenotyping is that new resistance conferring mutations can be discovered rapidly and in "real time" without the need for phenotyping to train the genotype. Moreover, in silico phenotypes are not subject to error caused from compensatory mutations which may act synergistically or anti-synergistically with resistance-associated mutations to increase, decrease, or reverse specific drug resistances. Computational phenotyping will generate information that can, for example, be presented in a report that is marketed within the in vitro diagnostics industry as an adjunct test/service to help optimize therapy and assist physicians, farmers, acadmenic institutions, government agencies, and industries with specimen treatment. Thus, a computer-based method for predicting clinical responses e.g. drug sensitivity or drug resistance in patients, plants, animals, pathogens, and microorganisms based on genetic polymorphisms is provided.

10

20

The genotypes used in the methods are obtained from any source, including, but are not limited to, from a plant, animal, pathogen, or mammal with the most preferred source being a mammal, paticularly a human for whom a particular drug treatment is contemplated, and is the genotype of the drug target, such as, as exemplified herein, HIV RT or PR from a particular infected individual. Other examplary drug targets are proteins, polypeptides, oligopeptides, including, but not limited to, a receptor, enzyme, hormone, and any such compound with which drugs or 25 other ligands interact to bring about a biological response. For exemplification of this method, the protein considered is an enzyme, in particular HIV protease (PR) and reverse transcriptase (RT), which are therapeutic drug targets. Nucleic acid encoding the target from individual sample, such as blood sample or other body fluid sample from a

-66-

mammal, such as a human patient, is sequenced, and the 3-D structure thereof determined. The drug of interest is computationally tested to assess whether it interacts with the sample.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLE 1

BINDING CORRELATIONS OF MUTANT FORMS OF HCV PROTEASE WITH DIFFERENT INHIBITORS

10 This example provides the results of a theoretical study of NS3 protease complexes with two known peptide inhibitors (see SEQ ID Nos. 1 and 2; Ingallinella et al. ((1998) Biochemistry 37:8906-8914).

Introduction

5

15

25

30

During HCV replication, the final steps of processing are performed by a virially encoded chymotrypsin-like serine protease NS3. NS3 is an approximately 3000 amino acid protein that contains, from the amino terminus to the carboxy terminus, a nucleocapsid protein (C), envelope proteins (E1 and E2) and several non-structural proteins (NS1, 2, 3, 4a, 4b, 5a and 5b). NS3 is an approximately 68 kDa protein, encoded by 20 approximately 1893 nucleotides of the HCV genome, and has two distinct domains: (a) a serine protease domain containing approximately 200 of the N-terminal amino acids; and (b) an RNA-dependent ATPase domain at the C-terminus of the protein. The NS3 protease is considered a member of the chymotrypsin family and is a serine protease that is responsible for proteolysis of the polypeptide (polyprotein) at the NS3/NS4a, NS4a/NS4b, NS4b/NS5a and NS5a/NS5b junctions responsible for generating four viral proteins during viral replication. This protease is inhibited by N-terminal cleavage products of substrate peptides. The NS3 protease, which is necessary for polypeptide processing and viral replication has been identified, cloned and expressed (see, e.g., U.S. Patent No. 5,712,145).

Active NS3 forms a heterodimer with a polypeptide cofactor NS4A. The crystal structure of NS3 with and without the NS4A cofactor is known (see, e.g., Love et al. (1996) Cell 87:331-342; Habuka et al. (1997) Jikken Igaku 15:2308-2313; Yan et al. (1998) Protein Sci. 7:837-847, which provides the structure with NS4A).

The NS3 protease is a target for design of antiviral drugs. For example, a series of potent hexapeptide inhibitors of NS3 has been developed by optimization of the product inhibitors (Ingallinella *et al.* (1998) *Biochemistry 37*:8906-8914).

10

25

_. . 5

Analyses

Models of the complexes of NS3 with the two protease inhibitor peptides were obtained by flexible docking of the peptides into the active site of the crystal structure of NS3/4A, followed by evaluation of protein-peptide binding energies. The models were tested by *in situ* modification of the docked ligands. A qualitative agreement between the binding energies and inhibitor IC_{50} values obtained from literature was found.

The peptides studied were:

	Sequence	IC ⁵⁰ , nM	SEQ ID
20	Ac-Asp¹-D-Glu²-Leu³-lle⁴-Cha⁵-Cys ⁶ -COO-	15	1
	Ac-Asp1-L-Glu2-Leu3-lle4-Cha5-Cys6-COO-	60	,2

• Cha = β -cyclohexylalanine

In the modeling studies, it was assumed that:

the high-affinity inhibitory peptides 1 and 2 have a similar mode of binding to the active site of NS3;

the minimum binding pharmacophore includes the SH group of Cys⁶ and carboxyl groups of Asp¹, Glu² and Cys⁶; and

the side chains of residues 3, 4 and 5 may enhance binding by non-specific hydrophobic interaction with NS3.

Methods

15

Initial structure of the NS3-peptide complex

The crystal structure of NS3 with a peptide cofactor NS4A was obtained from the arts (Kim et al. (1996) Cell 87:343) and was used in the studies with peptide inhibitors. The crystal structure of NS3/NS4A was regularized using molecular mechanics described herein. Initial NS3-NS4-peptide complexes were constructed by placing the peptides into the NS3 binding site expected by structural homology to by other serine proteases:

the C-terminal carboxyl was placed near the oxyanion-stabilizing site (residues 137-139);

the side chain of Cys⁶ was inserted into the hydrophobic cavity formed by L135, F154 and A157; and

the ϵ -amino group of K136 was placed in contact with the C-terminal carboxyl (see, Kim et al. (1996) Cell 87:343, Steinkuhler *et al.* (1998) *Biochemistry* 37:8899).

Monte Carlo simulations

In order to optimize the complexes, Biased Based Probability Monte Carlo (BPMC) simulations (Abagyan et al. (1994) J. Mol. Biol. 235:983)

20 were performed on the NS3-peptide complexes using the ICM program (commercially available from MolSoft, San Diego, CA) with ECEPP/3 force field and atomic solvation energies (Momany et al. (1975) J. Phys. Chem. 79:2361, Nemethy et al. (1992) J. Phys. Chem. 96:6472, Abagyan et al. (1997) Computer Simulations of Biomedical Systems: Theoretical and Experimental Applications, vol. 3, Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 363). The sampling method was BPMC with random change of one variable at a time. A Metropolis acceptance criterion was applied after energy minimization (quasi-Newton, up to 1000 steps). Simulations were performed at a temperature of 1000° K. The

5

10

15

20

25

peptide translational and rotational degrees of freedom, all peptide torsion angles and χ angles of the protein side-chains located within 7.0 Å of any peptide atom were varied during the BPMC simulations.

The energy function used in the MC simulations included:

ECEPP/3 terms for energy in vacuo (VDW (van der Waals), H-bond, electrostatic and torsion potentials);

distance dependent electrostatics with $e_0 = 4.0$; and surface energy with atomic solvation parameters.

The total energies of the complexes were calculated including contributions from: ECEPP/3 VDW, H-bond, S-S bond and torsion terms; exact-boundary electrostatic energy with $e_0 = 8.0$; and side-chain entropies. Hydrophobic free energies were estimated as sA, where A is accessible surface area and s is a tension constant of 0.03 kcal/molÅ².

Strategy of the flexible Monte Carlo docking

The simulations proceeded with multiple, relatively short MC runs (2000-5000 generated structures). New docking cycles were started from the lowest-energy or other interesting structures found in previous runs. Structures saved during various MC runs were sorted by total energies and RMSD (root-mean-squared deviation), and compressed into a cumulative conformational stack. Binding energies were calculated for representative structures of each complex thus obtained. This strategy was more efficient than continuous long simulations because the variable torsion angles and distance constraints are defined for an initial structure and do not change during the MC run.

Binding energies of the peptide-protein complexes

For low-energy conformations found after several iterative BMPC cycles, peptide-protein binding energies were estimated using the equation:

$$E_{bind} = E_o + E_{compl} - E_{pept} - E_{prot}$$

-70-

where E_{compl} is the energy of the complex, E_{pept} & E_{prot} are separate energies of the peptide and protein, respectively, and E_o is an adjustable constant.

The binding energy function included: exact-boundary electrostatic free energy contributions; side-chain entropy; and surface tension hydrophobic free energy terms. (Zhou and Abagyan (1998) Folding Design 3:513, Schapira *et al.* (1999) J. Mol. Recognition 12:177). ECEPP/3 hydrogen-bonding terms were included with a weight of 0.5.

Results

Models of the NS3-peptide complexes

RMSD between pharmacophore atoms of peptides 1 and 2 were calculated for all pairs of BPMC structures. Two models of the NS3-peptide complexes were selected assuming (1) similar positions of pharmacophore groups of two peptides in the binding site (RMSD \leq 2.0 Å) and (2) low binding energy of the complexes ($\Delta E_{bind} < 5.0$ kcal/mol). Two models of the NS3-peptide complex were selected by visual inspection.

Characteristics of the binding sites for peptide inhibitors in two NS3-peptide complex models are summarized in **Table 1**.

Table 1

Peptide site NS3 residue, group Type of Present for Peptide residue Model 2 interaction Model 1 P1 Cys⁶COO⁻ $K136 NH_3 +$ H-bond/el. 1,2 1,2 G137 NH 1,2 2 H-bond S139 OH H-bond 1,2 2 Cys⁶ SH L135, F154, A157 1,2 1,2 hydroph P2 Cha⁵ H57, R155, A156 hydroph 1,2 A157, V158 2 hydroph **P3** lle⁴ V132, S133 2 hydroph 1,2 V158, C159 hydroph 1

20

10

15

5

15

25

P4	Leu ³	Res. 157 to 160 V132, S133	hydroph hydroph	1,2	2
P5	Glu ² COO-	R161 guanidine	H-bond/el.		1,2
P6	Asp¹ COO-	R161 guanidine S133 OH	H-bond/el. H-bond	1,2	- 1,2

Validation of the models: modifications of the protein and ligands in the binding site

In order to validate the proposed models, the K136M mutation and peptide modifications known from SAR (structure-activity relationship) studies were performed in low-energy structures of the NS3-peptide 2 complex.

Positions of the modified ligand and conformations of adjacent protein side chains were adjusted by energy minimization. Distance restraints were applied to keep the ligand near its initial position.

Changes in calculated binding energies upon modifications, ΔE_{bind} (calc), were compared to the values expected from ratios of inhibitory potencies, ΔE_{bind} (exp).

$$\Delta E_{bind}(exp) = RT \ln(IC_{50}^{mod}/IC_{50}^{o}),$$

where IC_{50}° and IC_{50}^{mod} are inhibitory potencies of the parent and modified compounds.

The correlation between experimental and calculated changes in binding energy upon ligand modifications in the binding site of NS3 is illustrated in

FIG. 4.

Discussion

The two NS3-peptide complex models suggest a common binding pattern for the inhibitor P1 site (Cys⁶-OH) with the carboxyl group hydrogen-bonded to the oxyanion hole residues G137 and S139, and the

-72-

Cys⁶ side chain embedded in a hydrophobic pocket formed by L135, F154 and A157.

This study confirms the possibility of hydrogen bonding between the C-terminal carboxyl and ϵ -amino group of K136 suggested by Steinkuhler *et al.* ((1998) *Biochemistry* 37:8899) based on the K136M mutation in NS3. Changes in calculated binding energies upon mutation are consistent with an 8-fold increase in K_1 of an inhibitor with a free carboxyl group and with the lack of an effect on binding when the peptide is amidated.

The models differ in binding of the negatively charged side chains in positions P5 and P6. The R161 guanidine interacts with a carboxyl group of Asp¹ and Glu² in Models 1 and 2, respectively. In Model 2, the Asp¹ carboxyl also interacts with the hydroxyl of S133.

The models are in agreement with SAR data for peptide inhibitors of NS3. Predicted changes in binding energy upon modification of the protein and peptides correlate reasonably well with the changes expected from IC⁵⁰ ratios. Standard deviations of $\Delta E_{bind}(calc)$ - $\Delta E_{bind}(exp)$ were 0.8 and 1.6 kcal/mol for Models 1 and 2, respectively, with correlation coefficients of 0.62. After the largest outlier was removed from each dataset, correlations improved to 0.81 and 0.76, respectively.

Conclusions

10

15

20

25

An effective iterative Biased Probability Monte Carlo protocol for the docking of flexible peptide ligands into a flexible protein active site has been developed. Two models of the complexes of HCV NS3 protease with potent peptide inhibitors were proposed based on the docking simulations and on evaluation of protein-ligand binding energies. The models were validated by *in situ* modifications of NS3-peptide complexes and by correlation of binding energies of modified complexes with those expected from experimental IC₅₀ values. Proposed models can be used

10

20

25

30

for planning further mutagenesis studies of the HCV NS3 protease and the models can be used in the design of non-peptide inhibitors using structure-based drug design methodologies.

EXAMPLE 2

LEAD OPTIMIZATION BY RECEPTOR-BASED FREE ENERGY QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIPS (QSARS) FOR TNF RECEPTOR ANTAGONIST DISCOVERY

The goal of the modeling studies in this phase was to identify binding modes and complex structures of the compounds that bind to TNF receptor type I protein in order to guide the design of new compounds. An approach that relies on docking compounds to the receptor, evaluating free energy changes of binding of the docked structures, and comparing the calculated values with experimental inhibition constants K_i of the compounds was developed. The success of the calculations was assessed by evaluating the consistency of the calculated free energy changes of binding and the experimental K_i .

The difference in free energy changes of binding between two compounds with inhibition constants K_i and K_i can be calculated as,

 $\Delta\Delta G = -kT \ln K_i'/K_i$

where k and T are Boltzmann's constant and absolute temperature, respectively.

The 13 active compounds were studied. Their potencies, as measured by K_i , range from 0.1 to 30 μ M, spanning about 3 kcal/mol in free energy. It was found that the calculated free energy changes of binding are highly consistent with the corresponding experimental values, with correlation coefficient 0.966 and difference less than 0.5 kcal/mol (see Table 2 and Figure 4). The predicted binding modes and complex structures can thus be accepted with confidence.

To modify these compounds, important pharmacophore features on the surface of the receptor that are critical for binding of the compounds

were identified. These features include a hydrophobic belt, a hydrophilic belt and 3 hydrogen bond donor sites. A few of potential hydrogen bonding sites, which are not used by the current compounds, were also derived, and can be used for designing more potent binders.

Graphics-guided redesign of the compounds was performed. The free energy calculation was used to predict the binding activity of each design. Fourteen new compounds were thus designed and binding activities were predicted. The chemical structures of the designed molecules, together with the binding modes of the lead compounds, were synthesized and shown to have high affinity for the target. Some of them exhibit a K_i in low-nanomolar range. Hence the method provided herein for modification of drugs for binding to calculated 3-D structures of a target protein resulted in redesigned drug candidates with enhanced affinity for the target.

This approach has advantages over the traditional x-ray crystallography method, which include the following:

5

10

15

20

25

- (1) The binding modes are determined for a group of compounds instead of single compound; analysis of similarity and differences reveals rich information in binding mechanisms.
- (2) The predictive power of the free energy calculation is very desirable for redesign of compounds.
- (3) The correlation with the biochemical activities assures relevancy of the explored binding modes, while a structure given by x-ray crystallography may not necessarily be one related to the biological functions of the compound.

A comparison of calculated relative free energy changes of binding $\Delta\Delta$ A and experimental $\Delta\Delta$ G converted from inhibition constants K_i (all in kcal/mol) of the compounds (referenced by a code name) is presented in Table 2.

PCT/US00/30863 WO 01/35316

-75-

Table 2

	Compound	ΔΔΑ	ΔΔG
	SBI-2030	0	0
	SBI-2002	-0.97	-1.25
5	SBI-2005	-0.72	-1.14
	SBI-307	-0.56	-0.08
	SBI-2008	-0.53	-0.82
	SBI-2006	-0.34	-0.44
10	SBI-306	-0.07	0.40
	SBI-2000	0.29	0.27
	SBI-2001	0.72	1.12
	SBI-304	1.55	1.45
	SBI-308	1.70	1.78
	SBI-305	1.86	1.67
15	SBI-2048	1.95	1.94

A comparison of calculated versus experimental binding free energy changes is given in FIG. 5.

EXAMPLE 3

HIV Protease Models for Drug Studies 20

Antiviral therapy for AIDS has focused on the discovery and design of inhibitors for two main enzyme targets of the HIV-1: reverse transcriptase (RT) and protease (PR). HIV RT is a heterodimer composed of p51 and p66 subunits. The p51 subunit is composed of the first 450 amino acids 25 encoded by the RT gene and the p66 subunit is composed of all 560 amino acids of the RT gene. RT is responsible for RNA-dependent DNA polymerization, RNaseH activity, and DNA-dependent DNA polymerization.

HIV PR is a homodimer of two identical 99-amino acid chains. HIV PR is an aspartic proteinase that is responsible for the post-translational processing of the viral gag and gag-pol polyprotein gene products, which yields the structural proteins and enzymes of the viral particle (see, e.g., Erickson et al. (1996) Annu. Rev. Pharmacol. Toxicol. 36:545-571, Bouras et al. (1999) J. Med. Chem. 42:957-962). Despite several promising new anti-HIV agents, the clinical emergence of drug-resistant variants of HIV limits the long-term effectiveness of these drugs. Genetic analysis of the resistant forms of HIV has identified a number of critical mutations in the RT and PR genes. Moreover, structural analysis of inhibitor-enzyme complexes and mutational modeling studies can lead to a better understanding of how these drug-resistant mutations exert their effects at the structural and functional levels.

HIV-PR inhibitor computational binding studies

15 This example provides the results of a computational study on HIV PR. The 3-D protease structure was generated, docked with known viral inhibitors, and analyzed via free energy of binding studies described herein. A quantitative agreement between the calculated add experimental protease-drug binding energies was obtained. Moreover, a series of 3-D HIV PR models were analyzed to identify the invariant regions of the protease. These insights have implications for the design of new drugs and therapeutic strategies to combat AIDS drug resistance.

Optimization of 3D structures

20

Five PR inhibitors approved by the FDA for clinical use were used: 25 saquinavir, nelfinavir, indinavir, amprenavir, and ritonavir (Figure 6). Initial 3-D structures for the wild-type HIV PR complexes with these FDA approved inhibitors were obtained from the Protein Data Bank and were then optimized using Monte Carlo (MC) simulations with an ECEPP/3 force field as described in Example 1. The energy function used in the

25

MC simulations included: ECEPP/3 terms for energy in vacuo (van der

Waals, H-bond, electrostatic and torsion potentials); distance dependent dielectrics with e₀ = 4.0; and surface free energy calculated using atomic solvation parameters ((Dudek *et al.* (1998) *J. Computational Chem.* 19:548-573, Wang *et al.* (1995) J. Mol. Biol. 253:473-492). Standard ECEPP charges were used for the protein residues. Lys, Arg, Glu, and Asp residues were charged. Charged and protonated states of Asp 125 (chain B) were considered as well. The inhibitors were docked into the active site of the protease, and the protein-drug complexes were energetically refined using the methods described in Example 1. Partial charges for the inhibitors were calculated with the Gasteiger-Marsili method implemented in SYBYL 6.5 (Tripos Assoc., Inc.). Different protonation states were examined for indinavir and amprenavir, but the other inhibitors were assumed to be electroneutral. Water molecules located within 7.0 Å from a ligand atom in the X-ray structure were retained in the model complex during optimization.

Calculation of binding energies

For low energy conformations found after several iterative BMPC cycles, protein-drug binding energies were estimated using the equation:

 $E_{bind} = E_o + E_{compl} - E_{ligand} - E_{prot},$ where E_{compl} is the energy of the complex, E_{ligand} & E_{prot} are energies of the ligand and protein when separated, and E_o is an adjustable constant. The binding energies of the protein and ligand were calculated using the following energy function:

 $E=E_{el}+E_{vw}+E_{hb}+E_{s},$ where E_{el} is the exact-boundary electrostatic using $e_{0}=8.0$, E_{s} is the side-chain entropy term, and E_{vw} and E_{hb} are the ECEPP/3 van der Waals and hydrogen-bonding terms.

-78-

After the energies of the wild type PR-inhibitor complexes were calculated, mutation sites were introduced into the optimized X-ray structures or model complexes. The amino acid substitutions were followed by local optimization, using an ECEPP/3 force field, of protein side chains around the mutation sites via the energy minimization of substructures that included the ligand, water molecules within the sphere of radius 7.0 Å around the ligand, and protease residues within the sphere of radius 3-5 Å around the mutated residues. The energy of binding of the mutated complex was calculated based on the equation described herein. The difference in binding energy resulting from mutations (mut) of the wild-type (WT) protease were calculated using the following equation:

 ΔE_{bind} (calculated) = E_{bind} (WT) - E_{bind} (mut).

This change in binding energy was compared to data from experimental (exptl) studies (Gulnik et al. (1995) Biochemistry 35:9282-9287, Klabe et al. (1998) Biochemistry 37:8735-8742, Pazhanisami et al. (1996) J. Biol. Chem. 271:17979-17985, Jacobsen et al. (1995) Virology 206:527-534, Maschera et al. (1996) J. Biol. Chem. 217:33231-33235) based on the equation:

20 $\Delta E_{bind}(exptI) = RTln(K_imut/K_iwt)$.

10

15

25

Plots of ΔE_{bind} (calculated) vs. ΔE_{bind} (exptl) were generated, and the results, summarized in Table 3, show a strong correlation between the calculated binding energies and the experimentally determined binding energies for the PR-inhibitor complexes. For example, the correlation coefficient R for PR-ritonavir and PR-amprenavir is 0.9, where R=1 denotes congruency between the computationally calculated and experimentally determined binding energy data. These correlation data validate the computational protocol and calculations described herein as a method for predicting protein-drug binding or protein-drug resistance (i.e. non-binding). The

evaluation of changes in binding energy of protein-drug complexes upon protein sequence variations can be used as a possible descriptor and, thus, can be used to predict the efficacy of drugs on proteins resulting polymorphisms in genes. Moreover, the analysis of the free energy of binding in complexes between the protein models that are produced by the method set forth in this example and drugs that have been designed or modified is a good predictive tool for drug designers.

TABLE 3

Correlation between Experimental and Calculated Binding Energies
for HIV Protease Inhibitors

10

15

20

30

	101	IIIV I TOTE ASC T		
HIV PRInhibitor	X-ray Complex ID	No of exptl. data points	Correlation coefficient R	Correlation S.D., kcal/mol
Saquinavir	1HXB	18	0.84	0.68
Indinavir	1HSG	17	0.79	0.80
Ritonavir	1HXW	18	0.90	0.72
Amprenavir	1HPV	15	0.90	0.54
Nelfinavir	10HR	Insufficient data		

Identification of structural invariant regions of HIV Protease

Clinical effectiveness of HIV PR inhibitors is limited by the rapid emergence of drug-resistant mutations. Resistant PR variants first occur by the mutation of amino acids close to or in and around the drug binding site, which are then accompanied by compensatory mutations of more distant amino acids. The identification of highly conserved, structural invariant regions of a PR would provide new potential targets and thus lead to the development of therapeutics having greater clinical efficacy than those drugs commonly employed to treat HIV.

The protein sequences of HIV protease were obtained from GenBank and from the blood samples of patients using standard isolation and sequencing techniques well known in the arts. The protein sequences were modeled into 3-D structures using the computational

10

15

20

25

protocol described in Example 1. The protease sequences were aligned, and the frequency of mutation, regardless of type, was determined at each amino acid position and plotted in Figure 7, where the frequency of mutation in this set of HIV-1 Protease sequences varied from 0 to 40%. Sequence alignment also revealed how many different types of amino acids could be substituted in any specific residue, yielding the tolerance of each residue to substitutions of different types. The data showing the frequency of mutation of each residue out of PR sequences, the types of mutations, and the distance of the mutating residue from the active site (Asp 28) are shown in FIG. 8. This information, sequences obtained from 10591 different genotypes, was used to identify invariant and/or highly conserved regions of PR and to map these regions to a 3-D structure for the purpose of identifying new potential regions on the protein as targets for therapeutic intervention. These invariant regions include, but are not limited to, residues 1-9, 25-29, 49-52, 78-81, and 94-99, where residue 1 is an aliphatic amino acid, more preferably proline; residue 2 is a hydrophilic amino acid, more preferably glutamine; residue 3 is an aliphatic amino acid, more preferably isoleucine; residue 4 is a hydrophilic amino acid, more preferably threonine; residue 5 is a hydrophobic amino acid, more preferably leucine; residue 6 is an aromatic amino acid, more preferably tryptophan; residue 7 is a hydrophilic amino acid, more preferably glutamine; residue 8 basic amino acid, more preferably arginine; residue 9 is an aliphatic amino acid, more preferably proline; residue 25 is a hydrophilic amino acid, more preferably aspartic acid; residue 26 is a hydrophilic amino acid, more preferably threonine; residue 27 is an aliphatic amino acid, more preferably glycine; residue 28 is an aliphatic amino acid, more preferably alanine; residue 29 is an acidic amino acid, more preferably aspartic acid; residue 49 is an aliphatic amino acid, more preferably glycine; residue 50 is a hydrophobic amino acid,

more preferably isoleucine; residue 51 is an aliphatic amino acid, more preferably glycine; residue 52 is an aliphatic amino acid, more preferably glycine; residue 78 is an aliphatic amino acid, more preferably glycine; residue 79 is an aliphatic amino acid, more preferably proline; residue 80 is a hydrophilic amino acid, more preferably threonine; residue 81 is an aliphatic amino acid, more preferably proline; residue 94 is an aliphatic amino acid, more preferably glycine; residue 95 is a thio-containing amino acid, more preferably cysteine; residue 96 is hydrophilic amino acid, more preferably threonine; residue 97 is hydrophobic amino acid, more preferably leucine; residue 98 is hydrophilic amino acid, more preferably asparagine; and residue 99 is an aromatic amino acid, more preferably phenylalanine. These invariant regions can subsequently be used to assist in the design drugs or therapeutic agents which bind to the invariant regions and disrupt the activity of the protease with greater efficacy than drugs commonly used to treat HIV and where the free energy of binding between said drug or therapeutic agent and the structural invariant region is evaluated as described herein. The methods described in this example can also be applied to HIV RT and to any protein of interest that exhibits polymorphisms.

20 EXAMPLE 4

10

15

25

Computational Phenotyping of HIV-1 Protease and Reverse Transcriptase

Computational or *in silico* phenotyping is performed to assess phenotypic properties of a protein. This example demosntrates application of this method to HIV-1 protease and reverse transcriptase to test whether the efficacy of various protease inhibitors for an HIV patient.

To practice this method 3-D structures of HIV-1 protease and reverse transcriptase based upon the nucleic acid isolated from HIV from a patient are generated. Protein-drug binding analysis *in silico* in order to

determine whether drug binding does (i.e. sensitivity) or does not (i.e. resistance) take place.

Sequencing of HIV-1 Protease and Reverse Transcriptase is performed on HIV-1 cDNA following extraction, reverse transcription, and PCR amplification of viral RNA obtained from patient specimens, such as blood samples or other body fluid or tissue samples. Methods for the extraction, reverse transcription, and PCR amplification of viral RNA are well known in the art. For each sequence, a computer-generated 3-D structure of the protein is modeled and then docked with antiviral drugs in silico using methods described in Example 1 and elsewhere herein to analyze protein-drug interactions. Antiviral drugs that can be tested include, but are not limited to, saquinavir, indinavir, ritonavir, amprenavir, and nelfinavir for HIV protease; zidovudine, lamivudine, stavudine, zalcitabine, didanosine, abacavir, adefovir, delavirdine, nevirapine, and efavirenz for HIV reverse transcriptase; and any FDA-approved or non-FDA approved antiviral drug. From these protein-drug interaction studies, relative drug resistance or sensitivity is inferred by calculating and evaluating the free energy of binding in low energy conformations of complexes between the variant protease structure and docked antiviral drug or variant reverse transcriptase structure and docked antiviral drug, using the methods described in Examples 1 and 3 and elsewhere herein.

10

15

20

25

The results of the computational phenotyping procedure can be presented as a patient report that states whether a drug or drugs are sensitive or resistant to the RT or PR obtained from the patient. Such a patient report assists physicians in selecting appropriate drugs for HIV patients. It also is useful for the *in vitro* diagnostics industry in an adjunct test/service capacity to help optimize antiviral therapy.

-83-

EXAMPLE 5

HIV Protease and Reverse Transcriptase Databases

Exemplary databases of the 3-D protein structures of polymorphic variants are described in this example. The HIV PR and RT databases are a comprehensive collection of 3-D polymorphic structural data along with related information, including nucleic acids encoding all or a portion of the protein. These data provide a means to understand differences in the interactions between a drug or drugs and the structural variations of the drug targets.

This example describes the creation, interface for, and use of structural variant databases of HIV protease and reverse transcriptase polymorphic variants.

Construction of databases

20

25

To implement the RT or HIV database described herein, suitable computer for performing database server tasks includes a "Pentium" level 15 CPU having at least 128 MB of memory, 30 GB of disk storage, and 256 MB of disk swap space for files. A recommended configuration for better computer performance would include, for example, a "Pentium III" processor at 700 MHz or faster, memory of 256 MB or greater, disk storage space of 50 GB or more, and swap space of 500 MB or more. A suitable configuration for performing user tasks as described above includes a "Pentium" level CPU having 128 MB memory, disk space of 240 MB with swap space of 256 MB, and an optional display circuit card supporting OpenGL and having 4 MB of memory. A recommended configuration for better performance would include, for example, a "Pentium III" processor at 500 MHz or faster, memory of 256 MB or greater, disk space of 500 MB or more, swap space of 500 MB or more, and an optional display card having 8 MB of memory or more, supporting resolution of 1024 x 768.

15

20

Preferably, the software used in the computing system described above includes, for the server machine, operating system software such as "Windows NT Server 4.0" from Microsoft Corporation, with Service Pack 5, Version 1280 (10 June 1999) or more recent, with database management server software such as "Oracle Server Standard Edition 8.1" from Oracle Corporation, or better. The software used in a preferred embodiment of the user machine includes operating system software such as "Windows NT Workstation 4.0" from Microsoft Corporation, with Service Pack 5, version 1280 (10 June 1999) or more recent, as well as "Oracle Client Standard Edition Version 8.1" or better. The client machine will also be compliant with the "Java" programming language (Java Runtime Environment 1.2.2). As will be known to those skilled in the art, other configurations may be suitable, depending on the applications being used and the computer performance desired.

Database Interface

The database interface was a Java-based interface with useful features. The database is interfaced to a molecular graphics package that includes 3-D visualization, including wire-frame representations; secondary structure ribbons; and solid surfaces, and structure analysis tools. The database also provides an interface to access all of the collected files from the same 3-D structure. The database interface also provides access to other databases, such as databases of chemical structures and public domain databases such as GenBank and the Protein Data Bank. The OpenGL and C++ module has real-time interaction with the sequence display and sequence analysis modules, such that highlighting residues in one display results in highlighting those same residues in other displays.

The relational database containing the protein information may be structured according to relational objects to facilitate the analysis and

-85-

computation processes described in the preceding examples. FIG. 10 is a graphical representation of the database objects for the system described herein. The database is organized by classes, each of which is characterized by data attributes and subclasses for the proteins.

FIG. 10 shows that the database design includes classes comprising Variant and related classes of Sample, Residue, Model, Resistance_Entry, and Protein. Other classes include Conformation, Residue_Conformation, Atom, Drug, Family, and Subfamily. These classes store attribute data values and specify class parameters and behaviors to provide the functionality described herein.

_... 5

10

20

25

For example, FIG. 10 shows that the Variant class stores parameters to specify a variant, including subclasses that specify a Variant ID, Sample ID, Protein ID, Name, and Sequence, where Variant ID is the identification number of the variant; Sample ID is the identification number of the sample from which HIV PR and RT were obtained; Protein ID is the identification number of the protein i.e. PR or RT; Name is the name of the variant distinguishing it from other variants encoded by the same DNA due to ambiguities in the nucleic acid sequence; and Sequence is the nucleotide or amino acid sequence. Similarly, FIG. 10 shows that the Sample class includes subclasses relating to a specific sample and which specify Sample_ID, Sample_Date, Sex, Ambiguity Number, Distance, Sequence Length, Sequence, Clade, and Region, where Sample ID is as defined herein; Sample_Date is the date the sample was obtained; Sex is the gender of the sample donor; Ambiguity_Number is fraction of ambiguous nucleotide positions; Distance is a normalized number the variation of an amino acid from the master clade; Sequence_Length is the length of the sequence; Sequence is as defined herein; Clade is the master sequence; and Region is the geographic location from which the sample was obtained. The Model

-- · 5

20

25

class includes subclasses comprising Model_ID, Model_Name, Variant_ID, and Drug_ID, where Model_ID is the identification number of the 3-D protein model; Model_Name is the name of the 3-D protein model; Variant_ID is as defined herein; and Drug_ID is the identification number of the drug i.e. antiviral drug. The atom class includes the subclasses comprising Atom_Name, Residue_Conformation_ID, X_Coordinate, Y_Coordinate, and Z_Coordinate, where Atom_Name is the name of atom in the 3-D protein structure; Residue_Conformation_ID is the identification number of the amino acid conformation in a 3-D structure; and 10 X_Coordinate, Y_Coordinate, and Z_Coordinate are the coordinates of the 3-D protein structure. The conformation class includes the subclasses comprising Conformation_ID, Model_ID, and Refinement_Level, where Conformation_ID is the identification number of a conformation of a 3-D structure; Model_ID is as defined herein, and Refinement_Level is the number of times the conformation was refined energetically. The drug class includes the subclasses comprising Drug_ID, Profile, Symbol, Name1, Name2, Company, and URL, where Drug_ID is as defined herein; Symbol is the FDA symbol for the drug; Name1 is the name of the drug, Name2 is an alternative name of the drug; Company is the company that makes the drug; and URL is the website address of the company that makes the drug. The residue_conformation class includes the subclasses comprising Residue_Conformation_ID, Conformation_ID, and Residue_ID, where Residue_Conformation_ID is as defined herein; Conformation_ID is as defined herein; and Residue_ID is the identification number of the amino acid. The Resistance_Entry class includes the subclasses comprising Resistance_Entry_ID, Profile, Protein_ID, Residual_Number, Amino_Acid, Weight, and Maximum_Weight, where Resistance_Entry_ID is; Protein_ID is as defined herein, Amino_Acid is the amino acid. The Family class includes the subclasses comprising Family_ID and

-87-

Family Name, where Family ID is the identification number of the protein family and Family Name is the name of the protein family. The SubFamily class includes the subclasses comprising SubFamily ID, SubFamily Name, and Family_ID, where SubFamily_ID is the identification number of the protein subfamily, SubFamily_Name is the name of the protein subfamily, and Family ID is as defined herein. The Protein class includes the subclasses comprising Protein ID, Protein_Name, Species, Multiple_Domain, Multiple Chain, and Wild Type, where Protein ID is as defined herein, Protein Name is the name of the protein i.e. RT or PR; Species is the species of the source of the protein i.e. humans; Multiple_Domain is the domain of the protein i.e p66 or p51 in the case of RT; Multiple Chain is the a or b chain in the dimers of RT and PR; and Wild Type is the wild-type protein sequence for RT and PR. The residue class includes the subclasses comprising Residue ID, Variant ID, Chain, Residue_Number, Insertion_Code, and Residue Code, where Residue ID is the identification number of the amino acid, Variant ID is as defined herein, Chain, Residue_Number is the numbering of an amino acid in a protein sequence, Insertion_Code is the identification number if different insertions occur in the amino acid sequence, and Residue_Code is the single letter or 3-letter code of an amino acid. Those skilled in the art will understand the database design exemplified in FIG. 10. It should be understood that other classes or parameters may be included, as selected by those skilled in the art, for the desired database design.

Database Content

5

10

15

20

25 The databases contain information on the variants of HIV PR and RT present in patient populations. The master amino acid sequence, nucleic acid sequence, and 3-D structure are obtained from GenBank; an exemplary master sequence is set forth in SEQ ID No. 118. Nucleotide sequences exhibiting polymorphisms and the corresponding structural

variant protein sequences are determined by isolating nucleic from viruses and viral nucleic acid obtained from the blood samples of patients throughout the US, as well as from other countries, using sequencing methods well known in the art. The sequences were inputted into the RT and PR databases. Exemplary of the nucleotide sequences and the encoded amino acids for HIV RT and PR in this data base are set forth in SEQ ID NOS. 3 to 117, where r is g or a; y is t/u or c; m is a or c; k is g or t/u; s is g or c; w is a or t/u; b is g or c or t/u; d is a or g or t/u; h is a or c or t/u; v is a or g or c; and n is a or g or c or t/u or unknown or other. The amino acid sequences of the wild type and structural variants are used to create 3-D protein structures which are deposited into the databases.

1. 3-D Protein Models

10

25

The structure of the wild-type or master sequence model of PR and RT were obtained from the crystal structures found in PDB. The initial 15 structure was refined energetically using BPMC with an ECEPP force field as described in Example 1. The quality of the model was assessed by calculating Normalized Residue Energies (NREs), where models with $e_{av} \ge$ 1.5 require further energetic refinement; and models with $e_{av} < 1.5$ were deposited into the database as described herein. The 3-D protein 20 structures of the variant sequences were generated by comparing these structures to the master sequence (see, e.g., SEQ ID No. 118; i.e., homology modeling) and energetically refining the models ab initio, using the same force field and BPMC procedure as the master sequence and applying the same quality control standard as described herein. Figure 11 is a tabulation of the 3-D coordinates of an exemplary HIV PR entry in a database that includes 3-D structures. For US purposes and where permitted, Tables 4 and 5 are provided electronically on CD ROM. These Tables house the coordinates that represent the 3-D protein structures of

proteins encoded by the nucleic acids set forth in SEQ. ID. NOS. 3-117. It will be noted that these sequences encode a full length PR and about 200 nucleotides the p51 subunit, which is the subunit of interest herein. To construct the full-length 3-D structure, the 3-D structure of each encoded portion of the p51 subunit was generated and then combined with the structure of the master sequence to produce a full-length structure.

These 3-D structures in the database can be selected and exported into computational docking programs for analyzing protein-drug interactions on known drugs, new drugs or modified drugs. The database 10 can be mined to find protein models that correspond to patients with a particular genetic polymorphism, patients with the most commonly occurring polymorphism, to a relevant patient subpopulation (e.g., gender, age, race, or other characteristic), to patients receiving a specific 15 treatment regimen, to patients exhibiting a particular clinical response, to structural invariants, or to other relevant criteria. Drugs can be docked into the active sites of PR and RT and subsequently energetically refined using an ECEPP force field and BPMC as described in Example 1. The quality control is that the protein-drug complex 20 represents a low energy conformation, which may take several iterative BMPC cycles. Then, the binding energies of the protein-drug complexes can be estimated using the methods of Example 1. Drug designers can modify the structures of drugs or design new drugs, using methods well known in the arts, to maximize 25 the drug binding to the models generated by this database.

2. Other Data

Each PR or RT nucleotide sequence in the database has associated with it an identification number, the nucleotide sequence length, the translated amino acid sequence (or sequences in cases of ambiguous

20

nucleotide positions), a 3-D structure for each amino acid sequence (from which a number of structurally related values are calculated), the genotyping date, the gender of the patient, the geographical location from which the sample was sent, the clade of the sequence, the fraction of ambiguous nucleotide positions, drug information, and other clinical information.

Database Usage

A query menu allows the user to retrieve data based on the various fields: sample ID, residue number (with or without specific amino acid mutation), date gender, geographic location, distance from the master sequence, and other useful queries. The set of sequences that satisfies the user's query are brought up in a sequence display module, which have variations from the master sequence indicated initially, although the sequences can be highlighted according to predicted resistance. This subset of sequences can be subjected to further analyses. For example, a histogram summarizing the number of mutations at each position in the subset can be generated. The 3-D structures for any of the variants in the database can be displayed and analyzed in the structure visualization module, allowing the user to compare the similarities and differences between 3-D structures by superimposing the 3-D structures. The user and also export these structures into programs for protein-binding studies as described herein. Thus, by mining the databases, a user will access 3-D structures and clinical and sample information that can be used in and correlated with protein-drug binding studies of HIV PR and RT.

25 Database Applications

The HIV PR and RT databases have many applications. The applications include, but are not limited to, any application and method provided herein, such as databases that assist in de novo drug design and drug binding calculations. In particular, the database can be used in the

-91-

design of 2nd and 3rd generation drugs to combat potential resistance to HIV therapy, and it can be used in the design of drugs that will impact a broad spectrum of the infected population. The databases provide the ability to design drugs that focus on the most highly conserved regions of a drug target and drugs that will avoid resistance to mutation. The database could be used to rank drug candidates by likely efficacy within a given subpopulation of patients (e.g. age, race, gender) in pre-clinical trials and to predict the most effective drug regimen to give a patient, and for designing clinical trials.

10

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

CLAIMS

5

20

1. A computer-based method of drug design based on genetic polymorphisms, comprising:

obtaining more than one amino acid sequence of target proteins that are the product of a gene exhibiting genetic polymorphisms, wherein the sequences represent different genetic polymorphisms;

generating 3-dimensional (3-D) protein structural variant models from the sequences; and

based upon the structures of the 3-D models, designing drug

candidates, modifying existing drugs, identifying potential drug
candidates or identifying modifications of existing drugs based on
predicted intermolecular interactions of the drug candidates or modified
drugs with the structural variants.

2. The method of claim 1, wherein the structure-based drug design method comprises:

computationally docking the drug candidate or modified drug molecules with the target protein structural variant models;

energetically refining the docked complexes;

determining the binding interactions between the drug candidate or modified drug molecules and the structural variants; and

designing and identifying drugs or modifications to existing drugs based on the binding interactions.

- 3. The method of claim 2 wherein the binding interactions are determined by:
- calculating the free energy of binding between the protein structural variant model and the docked molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

4. The method of claim 1 wherein:

15

20

25

after the protein structural variant models derived from a particular genetic polymorphism are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models, wherein

the conserved structural features are used as a basis for structurebased drug design studies.

- 5. The method of claim 4, wherein the conserved structural features are stretches of non-contiguous residues, wherein each stretch contains at least two amino acids.
- 10 6. The method of claim 5, wherein the protein is human immunodeficiency virus protease.
 - 7. The method of claim 6, wherein the conserved residues comprise residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99; and wherein:
 - residue 1 is an aliphatic amino acid; residue 2 is a hydrophilic amino acid; residue 3 is an aliphatic amino acid; residue 4 is a hydrophilic amino acid; residue 5 is a hydrophobic amino acid; residue 6 is an aromatic amino acid; residue 7 is a hydrophilic amino acid; residue 8 is a basic amino acid; residue 9 is an aliphatic amino acid; residue 25 is an acidic amino acid; residue 26 is a hydrophobic amino acid; residue 27 is an aliphatic amino acid; residue 28 is an aliphatic amino acid; residue 29 is an acidic amino acid; residue 49 is an aliphatic amino acid; residue 50 is a hydrophobic amino acid; residue 51 is an aliphatic amino acid; residue 52 is an aliphatic amino acid; residue 78 is an aliphatic amino acid; residue 79 is an aliphatic amino acid; residue 80 is a hydrophilic amino acid; residue 95 is a thio-containing amino acid; residue 96 is a hydrophilic amino acid; residue 97 is hydrophobic amino acid; residue 98 is hydrophilic amino acid; and residue 99 is an aromatic amino acid.

- 8. The method of claim 6, wherein the conserved residues comprise residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99; and wherein:
- residue 1 is proline; residue 2 is glutamine; residue 3 is isoleucine; residue 5 4 is threonine; residue 5 is leucine; residue 6 is tryptophan; residue 7 is glutamine; residue 8 is arginine; residue 9 is proline; residue 25 is aspartic acid; residue 26 is threonine; residue 27 is glycine; residue 28 is alanine; residue 29 is aspartic acid; residue 49 is glycine; residue 50 is isoleucine; residue 51 is glycine; residue 52 is glycine; residue 78 is glycine; residue 79 is proline; residue 80 is threonine; residue 81 is proline; residue 94 is glycine; residue 95 is cysteine; residue 96 is threonine; residue 97 is leucine; residue 98 is asparagine; and residue 99 is phenylalanine.
 - 9. The method of claim 6, wherein the HIV protease has the sequence of amino acids set forth in any of SEQ ID Nos. 3-74 and 77-117.
 - 10. The method of claim 9, wherein the residues comprise residues 1-9, 25-29, 49-52, 78-81 and 94-99.
 - 10. The method of claim 1, wherein the selected model structures represent the structural variants resulting from the most commonly occurring genetic polymorphisms.

20

- 11 The method of claim 1, wherein the selected model structures represent the structural variants resulting from genetic polymorphisms found in a selected patient subpopulation.
- 12. The method of claim 1 wherein the structural variant models25 are stored in a relational database, comprising:
 - 3-D molecular coordinates for the structural variants;
 - a molecular graphics interface for 3-D molecular structure visualization; computer functionality for protein sequence and structural analyses; and

database searching tools.

5

20

- 13. The method of claim 12, wherein the database further comprises one or more of observed clinical data associated with the genetic polymorphisms, subject medical history and subject history.
- 14. The method of claim 1, wherein: after generating the 3-D protein structural variant models, the method comprises:

computationally docking drug molecules with the target protein models; and

- energetically refining the docked complexes; and
 wherein the candidate drugs are specific for a protein with a
 selected polymorphism or specifically interact with all proteins exhibiting a
 polymorphism.
- 15. The method of claim 14, wherein the structure-based drug15 design method comprises:

computationally docking drug or potential new drug candidate molecules with the target protein structural variant models;

energetically refining the docked complexes;

determining the binding interactions between the drug or potential

new drug candidate molecules and the structural variants; and

designing potential new drugs or modifications to existing drugs based on the binding interactions.

- 16. The method of claim 15, wherein the binding interactions are determined by:
- calculating the free energy of binding between the protein structural variant model and the docked molecule; and decomposing the total free energy of binding based on the interacting residues in the protein active site.
 - 17. The method of claim 14, wherein:

after the protein structural variant models derived from a particular genetic polymorphism are generated, selected model structures are analyzed to determine common structural features that are conserved throughout the selected models; and

the conserved structural features are used as a basis for structure-based drug design studies.

- 18. The method of claim 17, wherein the selected model structures represent the structural variants resulting from the most commonly occurring genetic polymorphisms.
- 19. The method of claim 17, wherein the selected model structures represent the structural variants resulting from genetic polymorphisms found in a specific patient subpopulation.
 - 20. The method of claim 12, wherein the selected model structures represent structural variants derived from patients the receive a specific treatment regimen.
 - 21. The method of claim 12, wherein the selected model structures represent structural variants derived from patients that exhibit a particular clinical responses to a given drug.
- 22. The method of claim 12, wherein the selected model
 structures represent structural variants derived based on the duration of a particular drug treatment.
 - 23. The method of claim 12, wherein the structural variant models are stored in a relational database, comprising:
 - 3-D molecular coordinates for the structural variants;
- a molecular graphics interface for 3-D molecular structure visualization; and

functionality for protein sequence and structural analysis; and database searching tools.

-97-

- 24. The method of claim 12, wherein the database further comprises observed clinical data associated with the genetic polymorphisms, subject medical history and subject history.
- 25. A computer-based method of selecting drug therapies for
 patients based on genetic polymorphisms, comprising:

obtaining amino acid sequences of a target protein that is the product of a gene exhibiting genetic polymorphisms, wherein the sequences represent different genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

computationally docking drug molecules with the target protein models;

energetically refining the docked complexes;

15

determining the binding interactions between the drug or potential new drug candidate molecules and the models; and

selecting drug therapies based on the drug or drugs that have the most favorable binding interactions with the structural variant models.

- 26. The method of claim 25, wherein the binding interactions are determined by:
- calculating the free energy of binding between the protein structural variant and the docked drug molecule; and

decomposing the total free energy of binding based on the interacting residues in the protein active site.

- 27. The method of claim 1, further after generating the 3-D
 25 structural variant models, exporting some or all of them models into a program that computationally docks the models with test compounds to assess intermolecular interactions.
 - 28. A computer-based method for predicting clinical responses in patients based on genetic polymorphisms, comprising:

10

25

obtaining one or more amino acid sequences for a target protein that is the product of a gene exhibiting genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

building a relational database of protein structural variants derived based on genetic polymorphisms and observed clinical data associated with particular polymorphisms exhibited in the patients, wherein the database comprises:

3-D molecular coordinates for the structural variant models; a molecular graphics interface for 3-D molecular structure visualization;

computer functionality for protein sequence and structural analysis;

database searching tools; and

observed clinical data associated with the genetic polymorphisms, subject medical history and subject history associated with the genetic polymorphisms;

obtaining a target protein structural variant based on the same gene associated with a polymorphism in a patient;

generating a 3-D protein model based on the subject's gene sequence;

screening/comparing the 3-D model derived from the subject to the structures contained in the database by:

identifying structures in the database that are similar to the model derived from the subject; and

predicting a clinical outcome for the patient based on the clinical data associated with the identified structures.

29. A computer-based method for designing therapeutic agents that are active against biological targets that have become drug resistant due to genetic mutations, comprising:

obtaining a first 3-D protein structural variant model of a target

protein against which a given drug has biological activity;

generating a second 3-D protein structural variant model of the target in which genetic mutations have occurred and against which the same drug is no longer biologically active;

comparing the structures of the first and second model to identify structural differences; and

performing structure-based drug design calculations in order to identify new drugs or modifications to the existing drug to bring about biological activity against the second model.

30. A computer-based method for identifying compensatory mutations in a target protein, comprising:

obtaining the amino acid sequence of a target protein containing multiple amino acid mutations that is expressed in a patient, wherein the structure of a form of the target protein that responds to a particular drug, including the active site, has been structurally characterized;

generating a 3-D structural model of the mutated protein;

20

25

comparing the structure of the mutated protein with the form of the protein that responds to the drug to identify structural differences and/or similarities arising from the mutations;

comparing the biological activities of the drug against both the mutated protein and the form of the protein that responds to the drug to determine the effects of the mutations on drug response; and

identifying the mutations in the protein that affect biological activity based on the comparisons.

31. A method for creating a 3-D structural polymorphism relational database, comprising:

obtaining one or more amino acid sequences of a target protein that is the product of a gene exhibiting a genetic polymorphism, wherein sequences represent different genetic polymorphisms;

generating 3-D protein structural variant models from the sequences;

energetically refining the models;

evaluating the quality of the models;

- optionally obtaining associated clinical properties or data; and inputting the model and any associated properties and/or data into a relational database.
 - 32. The method of claim 31, wherein after energetically refining the models, the models are further refined.
- 15 33. The method of claim 31, wherein the database comprises amino sequences of two or more polymorphic variants.
 - 34. The method of claim 31, wherein the database comprises amino sequences of ten or more polymorphic variants.
- 35. The method of claim 31, wherein the database comprises amino sequences of about 100 or more polymorphic variants.
 - 36. The method of claim 31, wherein the database comprises amino sequences of about 1000 or more polymorphic variants.
 - 37. The method of claim 31, wherein the database comprises amino sequences of more than 8000 polymorphic variants.
 - 38. A database created by the method of claim 31.
 - 39. The database of claim 38, comprising variant 3-dimensional structures of a selected target.
 - 40. The database of claim 38 that comprises structures of proteases or polymerases.

- 41. The database of claim 38, wherein the proteases are viral proteases or polymerases.
- 42. The database of claim 38, wherein the viral proteases are human immunodeficiency virus proteases and the polymerase is a viral reverse transcriptase.
- 43. The method of claim 31, wherein quality is assessed by computing the normalized residue energies such that if e_{av} is ≥ 1.5 a model is further refined until e_{av} is < 1.5; if e_{av} is < 1.5 a model is deposited into the database.
- 10 44. The method of claim 1, wherein the target is an enzyme.
 - 45. The method of claim 44, wherein the enzyme is a protease or polymerase.
 - 46. The method of claim 45, wherein the polymerase is a reverse transcriptase.
- 15 47. The method of claim 44, wherein the target is a protein expressed by an infectious agent.
 - 48. The method of claim 44, wherein the target is enzyme expressed by a an infectious agent.
- 49. The method of claim 48, wherein the agent is a human 20 immunodeficiency virus (HIV).
 - 50. A computer system, comprising a database containing data representative of the three dimensional structure of polymorphic variants of a drug target.
- 51. The system of claim 50, wherein the target is a cell surface receptor or an enzyme.
 - 52. The system of claim 50, wherein the enzyme is a protease or a polymerase.
 - 53. A database, comprising:

sequences of nucleotides encoding a protein or portions thereof, wherein proteins comprise polymorphic variants; and the portions encode a domain of the protein that comprises a site in the protein that binds to a drug candidates; and

- the coordinates of 3-dimensional (3-D) structures of the encoded proteins or portions thereof.
 - 54. The database of claim 53 that is a relational database.
 - 55. The database of claim 53 that comprises at least 2 polymorphic variants and the corresponding 3-D structures.
 - 10 56. The database of claim 55 that comprises at more than 10, more than 100, more than 1000, more than 8000, or more than 10,000 polymorphic variants and the corresponding 3-D structures.
 - 57. The database of claim 53, wherein the protein is a receptor or enzyme from a eukaryotic or prokaryotic organism.
- 15 58. The database of claim 53, wherein the organism is a pathogen or a mammal.
 - 59. The database of claim 53, wherein the organism is a pathogen is a virus or bacterium and the mammal is a human.
- 60. The database of claim 53, wherein the protein is a protease or a reverse transcriptase.
 - 61. A database, comprising the sequences of nucleotides set forth in SEQ ID Nos. 3-117 that encode HIV protease or the portion of HIV reverse transcriptase set forth in each SEQ ID.
- 62. The database of claim 53, further comprising 3-D structural coordinates for a protein or portion thereof comprising sequences of amino acids encoded by each of SEQ ID Nos. 3-117.
 - 63. The database of claim 54, wherein the protein is HIV protease.

- 64. The database of claim 54, wherein the protein is HIV reverse transcriptase.
- 65. The method of claim 1, wherein the target protein is a eukaryotic or prokaryotic protein.
- 66. The method of claim 1, wherein the target protein is an animal protein, a plant protein or a protein from a pathogen.

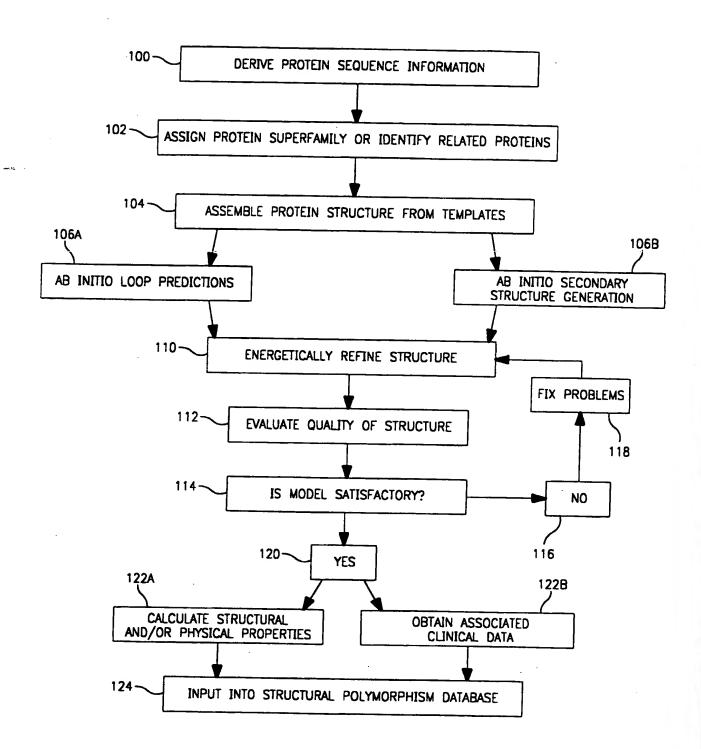


FIG. I

SUBSTITUTE SHEET (RULE 26)

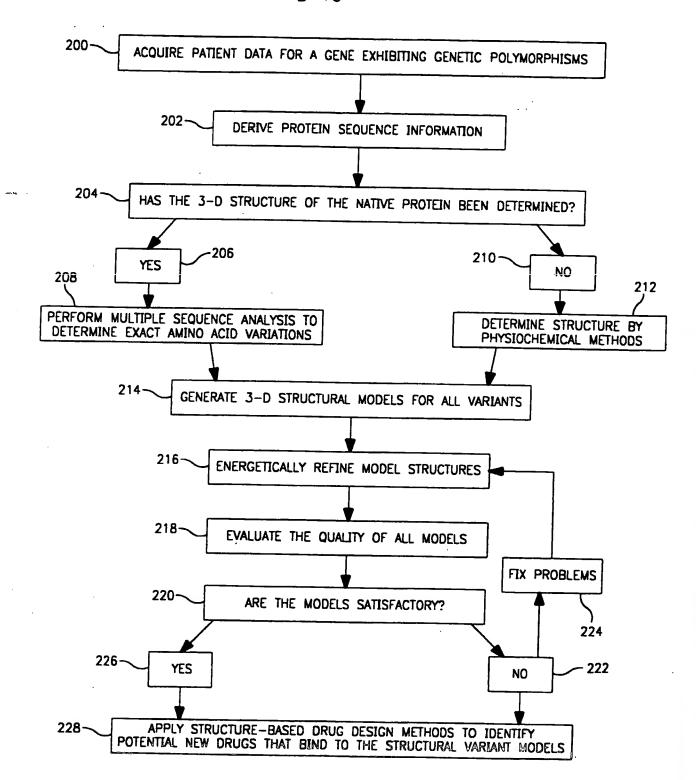


FIG. 2

SUBSTITUTE SHEET (RULE 26)

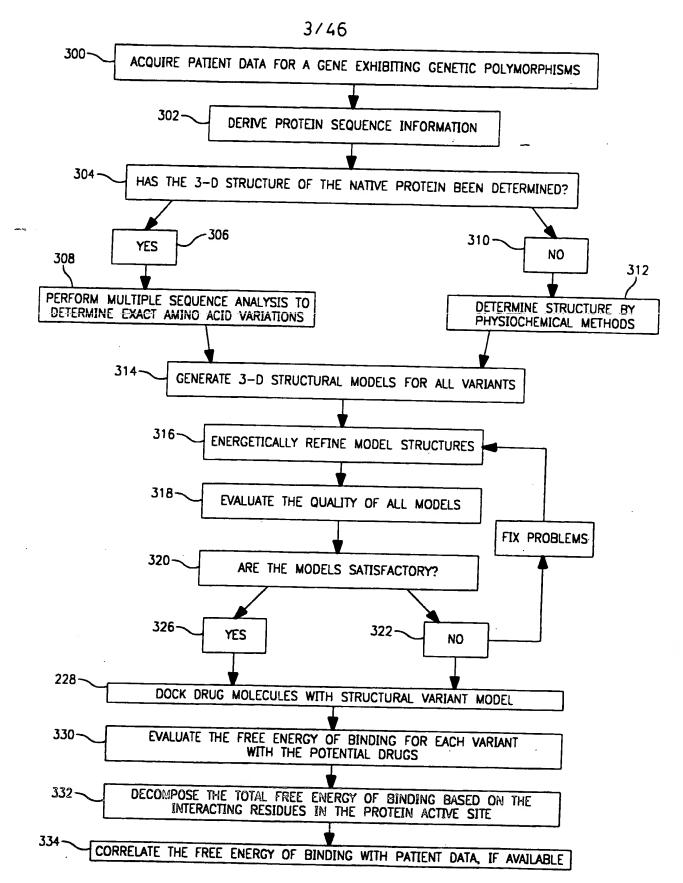
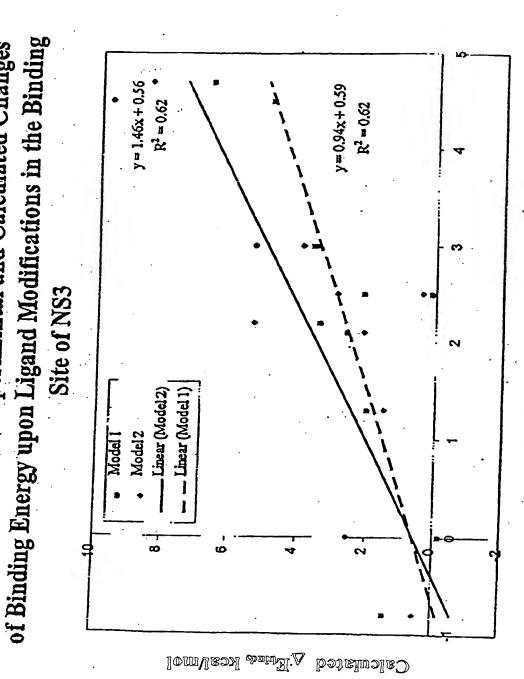


FIG. 3
SUBSTITUTE SHEET (RULE 26)

Correlation between Experimental and Calculated Changes F16. 🕶 4



Expected AEbus keal/mol

COMPARISON OF CALCULATED VERUS EXPERIMENTAL BINDING FREE ENERGY CHANGES

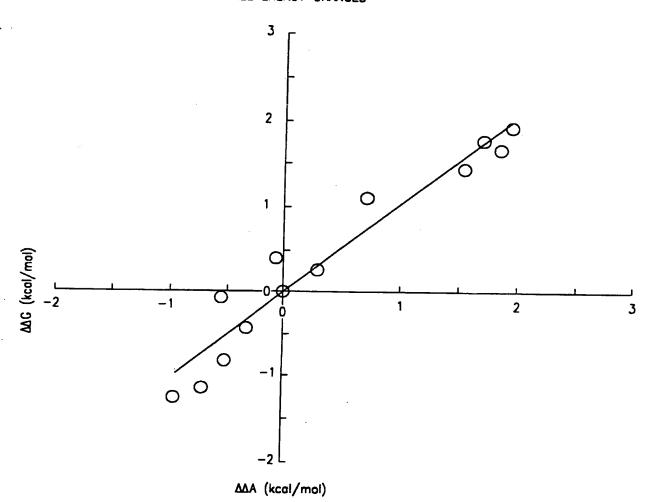
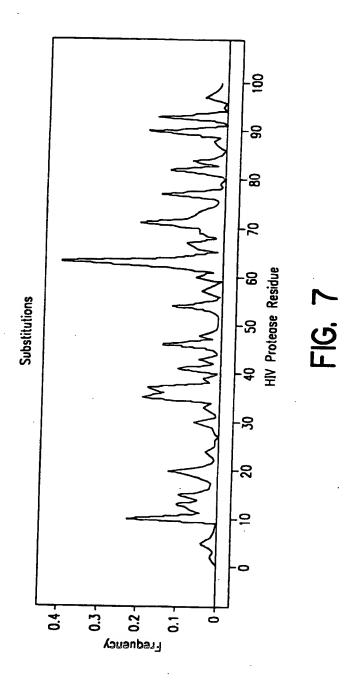


FIG. 5

HIV PROTEASE INHIBITORS APPROVED BY FDA

FIG. 6



SUBSTITUTE SHEET (RULE 26)

Database filename: hivpr.mdb Number of structures: 10591 Tolerance (%) >= 1.05

- (- / -							
ResNum	Tot0cc	TotFreq	Dist	WtAA	NumMut	MutList	NumList
1	11	0	15.4	P	0		
2	32	0	14.5	Q	0		
3	38	0	12.1	1	0		
4	106	0	13.0	T	0		
5	100	0	11.3	L	0		
6	47	0	14.3	W	0		
7	58	0	12.8	Q	0		
8	27	0	9.6	R	0		
9	11	0	7.9	P	0		
10	4004	37.8	9.2	L	3	IV F	3162 441 278
11	82	0	10.9	· . V	0		
12	1117	10.5	13.7	T	5	SEPAN	241 185 158 155 117
13	1745	16.5	13.7	I	1	V	1717
14	646	6.1	17.0	K	1	R	623
15	1760	16.6	17.5	I	1	V	1709
16	361	3.4	20.9	G	1	E	254
1.7	56	0	22.4	G	0.	•	
18	242	2.3	20.5	Q	0		
19	1340	12.7	18.3	L	4	IVQT	873 162 130 128
20	1549	14.6	15.4	K	4	IRTM	576 560 209 145
21	43	Ō	12.7	Ε	0		
22	46	0	9.0	A	0		
23	89	0	5.8	L	0		
24	402	3.8	3.8	Ĺ.	1	1	377
							- ·

FIG. 8A

25	28	0	0.0	D	0		
26	- 14	0	3.8	T	0		
27	9	0	5.5	G	0		
28	16	0	5.8	Α	Ō	•	
29	34	0	8.7	D	0		
30	770	7.3	9.2	D	1	N	725
31	15	0	8.9	T	0		
32	238	2.2	10.5	V	1	1	221
33	578	5.5	12.4	Ľ	3	VIF	207 189 172
34	88	0	15.1	Ε	0		100 172
35	2790	26.3	18.6	Ε	1	D	2646
36	2780	26.2	20.2	М	2	IV	2549 129
37	3252	30.7	22.8	N	4	DSET	1253 1129 246 209
38	54	0	22.0	L	0		
39	302	2.9	24.9	P	1	S	133
40	19	0	25.5	G	0		
41	2249	21.2	26.0	R	1	K	2235
42	21	0	23.5	w	0		
43	372	3.5	23.7	K	2	TR	166 144
44	12	0	22.6	P	0		
45	208	2	20.0	K	1	R	170
46	2165	20.4	18.8	M	2	IL	1580 560
. 47	47	. 0	15.4	1	0		
48	445	4.2	14.9	G	1	٧	385
49	17	0	12.9	G	0		
50	31	0	14.5	I	0		,
51	24	0	17.6	G	0		
52	12	. 0	18.3	G	0		
53	408	3.9	18.1	F	1	L	360



54	1661	15.7	18.0	1	1	V	1460
55	164	1.5	19.7	K	1	R	149
56	13	0	18.1	V	0		
57	1194	11.3	19.7	R	1	K	1162
58	341	3.2	18.6	Q	1	Ε	317
59	20	0	19.4	Y	0		
60	992	9.4	19.6	D	1	Ε	938
61	468	4.4	19.9	Q	1	Ε.	285
62	2711	25.6	18.6	j	1	٧	2685
63	8864	83.7	18.5	L	6	PASTQH	7245 380 321 266 226 162
64	2238	21.1	15.8	I	2	VL	1931 223
65	222	2.1	15.6	Ε	1	D	206
66	194	1.8	12.8	1	0		
67	309	2.9	14.6	С	1	S	143
68	51	0	17.5	G	0		
69	773	7.3	16.1	Н	2	QY	376 206
70	478	4.5	17.0	K	1	R	359
71	3664	34.6	15.3	A	3	VII	2301 1145 190
72	1494	14.1	17.2	1	3	VTL	650 409 126
73	1246	11.8	15.8	G	2	ST	932 185
74	658	6.2	15.4	T	2	SA	433 126
75	73	0	14.1	V	0		
76	59	0	14.6	L	0		
77	3533	33.4	16.1	V	1	1	3513
78	8	0	16.9	G	0		
79	95	0	17.2	P	0		
80	6	0	13.6	T	0		
81	7	0	13.7	P	0		
82	2208	20.8	11.0	V	2	AT	1668 284



83	44	0	9.7	N	0		
84	1091	10.3	6.3	I	1	V .	1073
85	213	2	5.7	Ţ	i	٧	198
86	16	0	5.3	G	0		
87	32	0	7.3	R	0		
88	706	6.7	10.4	. N	2	DS	543 128
89	240	2.3	10.1	L	1	M	143
90	3429	32.4	8.3	. L	1	M	3397
91	28	0	11.4	T	0		
92	227	2.1	13.6	Q	1	K	169
93	3095	29.5	13.1	I	1	L	3041
94	15	0	13.6	G	0		
95	100	0	10.6	С	0		
96	6	0	11.2	T	0		
97	83	0	10.7	L	0		
98	44	0	14.2	N	0		
99	35	0	16.4	F	0		

FIG. 8D

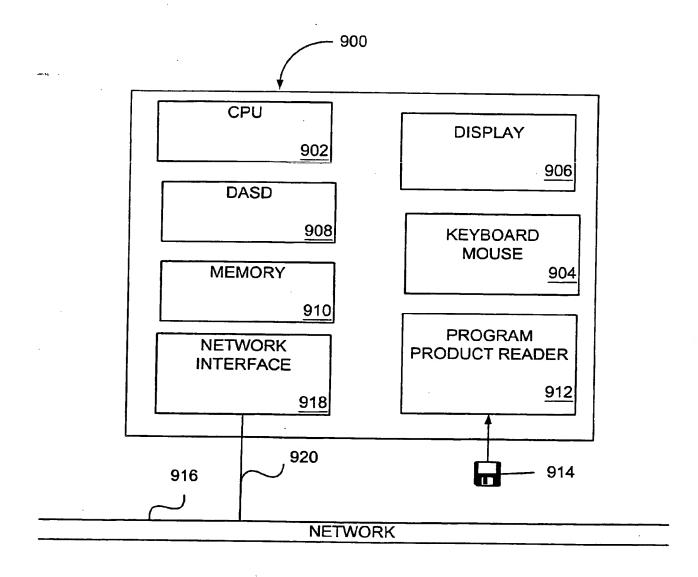
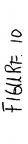
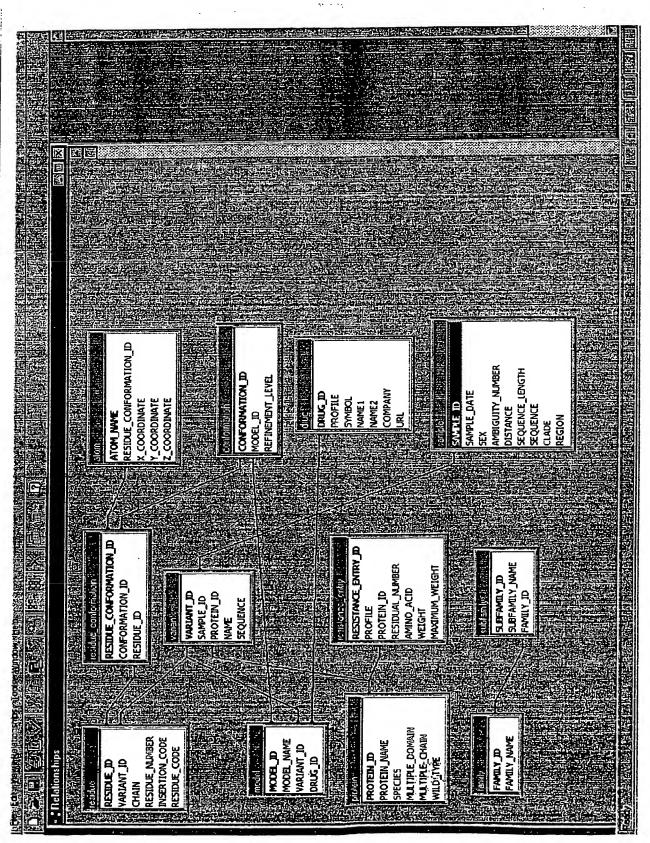


FIG. 9





WO 01/35316 PCT/US00/30863

					14	146			
N TOM		,	NŤ	PRO		1	-3.433	7.956	34.152
ATOM		1	N			1	-2.653	6.918	34.784
MOTA		2	CA	PRO			-1.242	7.005	34.259
ATOM		3	C	PRO		1	-0.950	7.638	33.216
ATOM		4	0	PRO	A	1	-3.281	5.601	34.262
MOTA		5	CB	PRO	A	1	-4.191	5.995	33.118
MOTA		6	CG	PRO	Α	1		7.461	33.339
MOTA		7	CD	PRO	Α	1	-4.547	8.493	33.547
MOTA		8	1H	PRO	Α	1	-2.845	8.552	34.853
MOTA		9	2H	PRO	Α	1	-3.824 -0.259	6.464	35.001
ATOM		10	N	GLN		2	-0.239	6.057	35.889
ATOM		11	H	GLN		2	1.115	6.443	34.568
MOTA		12	CA	GLN	A	2	1.452	4.993	34.301
MOTA		13	C	GLN		2	1.379	4.106	35.173
MOTA		14	0	GLN		2		6.966	35.653
ATOM		15	CB		A	2	2.070	6.859	35.240
MOTA		16	CG	GLN	A	2	3.549	7.744	36.054
ATOM		17	CD	GLN	Α	2	4.490	8.888	35.719
ATOM		18	0E1	GLN	A	2	4.771	7.190	37.144
ATOM		19	NE2	GLN	A	2	4.980	7.702	37.734
MOTA		20	1HE2	GLN	A	2	5.605 4.731	6.253	37.734
MOTA		21	2HE2		A	2		4.644	33.037
ATOM		22	N		A	3	1.784	5.351	32.336
MOTA		23	H		A	3	1.876 2.013	3.257	32.665
MOTA		24	CA	ILE	A	3	3.505	3.028	32.473
MOTA		25	C	ILE	A	3	4.242	3.777	31.787
ATOM		26	0	ILE	A	3	1.226	2.944	31.370
ATOM		27	CB		A	3. 3	-0.274	3.239	31.603
ATOM		28	CG1		A	3	1.427	1.480	30.901
ATOM		29	CG2		A A	3	-1.089	3.219	30.322
ATOM		30	CD1	THR		4	4.071	2.032	33.177
ATOM		31	N H	THR		4	3.525	1.525	33.844
ATOM ATOM		32 33	CA	THR		4	5.451	1.661	33.007
ATOM	•	34	C		A	4	5.515	0.637	31.901
ATOM		35	0	THR		4	4.490	0.143	31.397
ATOM		36	CB		A	4	6.051	1.125	34.324
ATOM		37	OG1		A	4	5.224	0.069	34.791
ATOM		38	HG1	THR		4	5.589	-0.299	35.646
ATOM		39	CG2	THR		4	6.085	2.212	35.431
ATOM		40	N	LEU		5	6.677	0.281	31.405
ATOM		41	H	LEU		5	7.518	0.530	31.885
ATOM		42	CA	LEU		5	6.754	-0.464	30.177
ATOM		43	C	LEU		5	7.432	-1.813	30.356
ATOM	•	44	Õ	LEU		5	7.940	-2.464	29.426
ATOM		45	СВ	LEU		5	7.459	0.394	29.128
ATOM		46	CG	LEU		5	6.668	1.671	28.775
ATOM		47	CD1	LEU		5	7.493	2.649	27.939
ATOM		48	CD2	LEU		5	5.345	1.307	28.099
ATOM		49	N		A	6	7.420	-2.351	31.594
ATOM		50	H		A	6	7.030	-1.833	32.356
ATOM		51	ĊA	TRP	A	6	7.958	-3.669	31.865
ATOM		52	C		A	6	7.071	-4.697	31.204
ATOM		53	ō	TRP		6	7.520	-5.798	30.828
ATOM		54	CB	TRP		6	8.099	-3.913	33.367
ATOM		55	CG	TRP		6	9.041	-2.974	34.070

FIG. I IA SUBSTITUTE SHEET (RULE 26)

				15	146	
ATOM	56	CD1	TRP	Α	6	8.745 -1.769 34.646
ATOM	57			A.	6	10.449 -3.171 34.273
ATOM	58	NE1		Α	6	9.875 -1.209 35.190
ATOM	59			Α	6	9.930 -0.332 35.668
ATOM	60			Α	6	10.932 -2.048 34.974
ATOM	61	CE3		Α	6	11.334 -4.190 33.924
ATOM	62	CZ2		Α	6	12.257 -1.917 - 35.333
ATOM	63	CZ3		Α	6	12.650 -4.065 34.278
ATOM	64	CH2	TRP	Α	6	13.106 -2.942 34.974
ATOM	65	N	GLN	Α	7	5.773 -4.448 30.973 5.354 -3.619 31.343
ATOM	66	H	GLN	Α	7	
ATOM	67	CA	GLN	Α	7	
ATOM	68	С	GLN	A	7	
ATOM	69	0	_	Α	7	
ATOM	70	CB		Α	7	3
ATOM	71	CG	-	A	7	
ATOM	72	CD		Α	7	
ATOM	73	OE1		Α	7	2.053 -7.681 32.712 2.356 -5.682 33.736
MOTA	74	NE2	GLN	A	7	1.501 -5.748 34.251
ATOM	75	1HE2	GLN	A	7	2.926 -4.867 33.837
ATOM	76	2HE2	GLN	A	7	3.777 -5.239 28.078
ATOM	77	N	ARG	Α	8.	3.688 -6.233 28.142
ATOM	78	H	ARG	A	8 8	3.183 -4.568 26.948
ATOM	79	CA		A	8	2.117 -3.648 27.461
ATOM	80	C		A A	8	1.333 -3.965 28.387
ATOM	81	O		A	8	2.574 -5.555 25.975
ATOM	82	CB	ARG ARG	A	8	3.532 -6.593 25.437
MOTA	83	CG CD	ARG	A	8	2.842 -7.610 24.579
MOTA	84 85	NE		A	8	3.787 -8.487 23.900
MOTA	86	HE	ARG	A	8	4.762 -8.279 23.982
MOTA MOTA	87	CZ	ARG	A	8	3.405 -9.541 23.185
ATOM	88	NH1	ARG	A	8	2.125 -9.871 23.052
ATOM	89	2HH1	ARG	Α	8	1.418 -9.321 23.496
ATOM	90	1HH1	ARG	Α	8	1.869 -10.670 22.508
ATOM	91	NH2	ARG	Α	8	4.332 -10.286 22.589
ATOM	92	1HH2	ARG	A	8	4.062 -11.082 22.048
ATOM	93	2HH2	ARG	Α	8	5.299 -10.050 22.682 1 990 -2.428 26.938
ATOM	94	N	PRO	Α	9	1.330
MOTA	95	CA	PRO		9	1.001 -1.462 27.440 -0.365 -1.697 26.821
MOTA	96	С	PRO		9	
MOTA	97	0	PRO		9	
ATOM	98	CB	PRO		9	
MOTA	99		PRO		9	2.334
ATOM	100		PRO		9	3.024 -1.820 26.084 -1.028 -2.803 27.227
MOTA	101	N	LEU		10	-0.616 -3.404 27.912
MOTA	102		LEU		10	-2.319 -3.143 26.698
ATOM	103		LEU		10	-3.390 -2.565 27.591
ATOM	104		LEU		10	-3.336 -2.632 28.831
ATOM	105		LEU		10 10	-2.451 -4.651 26.709
atom	106		LEU		10	-1.483 -5.316 25.756
ATOM	107				10	-1.159 -6.740 26.212
ATOM	108				10	-2.083 -5.262 24.322
MOTA	109		VAL		11	-4.447 -1.952 27.033
MOTA	110		VAL		11	-4.507 -1.875 26.038
MOTA	111	. п	AVT			

FIG. I IB

WO 01/35316 PCT/US00/30863

ATOM 112 CA VAL A 11					1	6/46			
ATOM 113 C VAL A 11	T TOM	112	CD	VAI.			-5.506	-1.398	
ATOM 114 O VAL A 11 -6.924 -2.490 26.198 ATOM 115 CB VAL A 11 -5.420 0.143 27.897 ATOM 116 CGI VAL A 11 -5.420 0.143 27.897 ATOM 117 CG2 VAL A 11 -5.549 0.787 26.497 ATOM 118 N THR A 12 -7.954 -1.592 -27.978 ATOM 119 H THR A 12 -7.984 -1.141 28.868 ATOM 119 H THR A 12 -7.984 -1.141 28.868 ATOM 120 CA THR A 12 -9.884 -1.141 28.868 ATOM 120 CA THR A 12 -9.889 -0.726 26.795 ATOM 122 O THR A 12 -9.889 -0.726 26.795 ATOM 123 CB THR A 12 -9.889 -0.726 26.795 ATOM 123 CB THR A 12 -9.596 -3.458 29.338 ATOM 124 OGI THR A 12 -9.596 -3.458 29.338 ATOM 125 HGI THR A 12 -10.225 -2.385 28.659 ATOM 126 CG2 THR A 12 -10.570 -3.766 00.096 ATOM 127 N ILE A 13 -10.449 -0.932 25.594 ATOM 128 H ILE A 13 -10.449 -0.932 25.594 ATOM 129 CA ILE A 13 -10.409 -1.841 25.178 ATOM 130 C ILE A 13 -10.409 -1.841 25.178 ATOM 131 O ILE A 13 -10.409 -1.841 25.178 ATOM 132 CB ILE A 13 -10.432 0.364 23.511 ATOM 133 CGI ILE A 13 -10.466 -0.896 22.628 ATOM 133 CGI ILE A 13 -10.466 -0.896 22.628 ATOM 134 CG2 ILE A 13 -10.466 -0.896 22.628 ATOM 135 CDI ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 136 CDI ILE A 13 -9.755 -0.745 21.294 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 136 CDI ILE A 13 -9.755 -0.745 21.294 ATOM 137 H LYS A 14 -13.809 1.622 24.481 ATOM 137 H LYS A 14 -13.809 1.622 24.481 ATOM 137 H LYS A 14 -13.809 1.622 24.481 ATOM 140 CDI LYS A 14 -14.859 2.059 2.375 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -19.816 0.929 2.375 ATOM 142 CG LYS A 14 -19.816 0.929 2.375 ATOM 145 NZ LYS A 14 -19.816 0.929 2.375 ATOM 145 NZ LYS A 14 -19.816 0.929 2.375 ATOM 145 NZ LYS A 14 -19.816 0.929 2.375 ATOM 145 NZ LYS A 14 -19.816 0.929 2.375 ATOM 148 2HZ LYS A 14 -19.816 0.929 2.375 ATOM 149 N ILE A 15 -15.642 0.347 0.065 20.091 ATOM 145 NZ LYS A 14 -19.816 0.929 2.375 ATOM 146 HZ LYS A 14 -19.816 0.929 2.375 ATOM 155 CGI ILE A 15 -15.642 0.347 0.065 20.091 ATOM 155 CGI ILE A 15 -15.642 0.347 0.040 0.328 19.887 ATOM 155 CGI ILE A 15 -15.642 0.347 0.040 0.328 19.								-1.857	27.268
ATOM 115 CB VAL A 11								-2.490	
ATOM 116 CG1 VAL A 11 -4.117 0.595 28.551 ATOM 117 CG2 VAL A 11 -5.549 0.787 26.4597 ATOM 118 N THR A 12 -7.954 -1.592 27.978 ATOM 119 H THR A 12 -7.884 -1.141 28.868 ATOM 120 CA THR A 12 -9.801 -1.942 27.496 ATOM 121 C THR A 12 -9.889 -0.726 26.795 ATOM 122 O THR A 12 -9.856 0.436 27.247 ATOM 123 CB THR A 12 -9.856 0.436 27.247 ATOM 124 OG1 THR A 12 -9.856 0.436 27.247 ATOM 125 HG1 THR A 12 -9.596 -3.458 29.338 ATOM 125 HG1 THR A 12 -9.596 -3.458 29.338 ATOM 126 CG2 THR A 12 -10.170 -3.766 30.096 ATOM 126 CG2 THR A 12 -10.170 -3.766 30.096 ATOM 127 N ILE A 13 -10.449 -0.932 25.594 ATOM 128 H ILE A 13 -10.449 -0.932 25.594 ATOM 129 CA ILE A 13 -10.449 -1.841 25.178 ATOM 130 C ILE A 13 -11.112 0.133 24.882 ATOM 131 O ILE A 13 -12.553 -0.292 24.693 ATOM 133 CG1 ILE A 13 -10.432 0.364 23.511 ATOM 133 CG1 ILE A 13 -10.432 0.364 23.511 ATOM 133 CG1 ILE A 13 -10.432 0.364 23.511 ATOM 135 CG1 ILE A 13 -10.4466 -0.896 22.628 ATOM 136 N LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -13.808 0.877 22.719 ATOM 140 O LYS A 14 -14.859 2.059 2.2375 ATOM 140 C LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -15.855 0.916 25.099 ATOM 145 NZ LYS A 14 -18.826 0.150 28.179 ATOM 146 HZ LYS A 14 -19.801 1.693 28.559 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATOM 150 C ILE A 15 -15.642 0.347 20.400 ATO									27.897
ATOM 118 CG2 VAL A 11									28.551
ATOM 118 N THR A 12									26.497
ATOM 119 H THR A 12 -7.884 -1.141 28.868 ATOM 120 CA THR A 12 -9.301 -1.942 27.496 ATOM 121 C THR A 12 -9.889 -0.726 26.795 ATOM 122 O THR A 12 -9.856 0.436 27.247 ATOM 123 CB THR A 12 -9.596 -3.458 29.338 ATOM 125 HG1 THR A 12 -10.170 -3.766 30.096 ATOM 126 CG2 THR A 12 -10.170 -3.766 30.096 ATOM 127 N ILE A 13 -10.449 -0.932 25.594 ATOM 128 H ILE A 13 -10.409 -1.841 25.178 ATOM 128 H ILE A 13 -10.409 -1.841 25.178 ATOM 131 O ILE A 13 -10.409 -1.841 25.178 ATOM 131 O ILE A 13 -10.409 -1.841 25.178 ATOM 133 CB ILE A 13 -10.409 -1.841 25.178 ATOM 134 CG2 ILE A 13 -10.432 0.364 23.511 ATOM 135 CD1 ILE A 13 -10.432 0.364 23.511 ATOM 135 CD1 ILE A 13 -10.466 -0.896 22.628 ATOM 136 N LYS A 14 -13.209 1.652 24.481 ATOM 137 H LYS A 14 -13.209 1.652 24.481 ATOM 138 CA LYS A 14 -13.209 1.652 24.481 ATOM 139 C ILYS A 14 -13.209 1.652 24.481 ATOM 139 C ILYS A 14 -13.209 1.652 24.481 ATOM 137 H LYS A 14 -13.209 1.652 24.481 ATOM 137 H LYS A 14 -13.209 1.652 24.481 ATOM 138 CA LYS A 14 -13.470 0.658 24.438 ATOM 140 O LYS A 14 -15.088 0.877 22.719 ATOM 141 CB LYS A 14 -15.088 0.877 22.719 ATOM 142 CG LYS A 14 -15.855 0.916 25.099 ATOM 144 CE LYS A 14 -18.678 0.162 24.481 ATOM 145 NZ LYS A 14 -18.678 0.166 25.099 ATOM 145 NZ LYS A 14 -18.678 0.166 25.099 ATOM 145 NZ LYS A 14 -18.678 0.166 25.099 ATOM 146 LYZ A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 146 LYZ A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.806 0.877 22.719 ATOM 148 2HZ LYS A 14 -19.806 0.877 22.719 ATOM 146 LYZ A 14 -19.806 0.806 23.747 20.400 ATOM 150 C ILE A 15 -15.662 0.347 20.400 ATOM 150 C ILE A 15 -15.662 0.347 20.400 ATOM 155 CGI ILE A 15 -15.662 0.347 20.400 ATOM 155 CGI ILE A 15 -15.662 0.347 20.400 ATOM 155 CGI ILE A 15 -15.662 0.347 20.400 ATOM 155 CGI ILE A 15 -16.894 -0.328 19.867 ATOM 156 CGI ILE A 15 -16.894 -0.328 19.867 ATOM 156 CGI ILE A 15 -16.894 -0.328 19.867 ATOM 156 CGI ILE A 15 -16.894 -0.328 19.867 ATOM 156 CGI ILE A 15 -16.897 -0.493 21.988 ATOM 156 CGI ILE A 15 -16.693 -0.493 21.988 ATOM 156 CGI ILE								-1.592	.27.978
ATOM 120 CA THR A 12									28.868
ATOM 121 C THR A 12									
ATOM 122 O THR A 12									
ATOM 123 CB THR A 12									
ATOM 124 OG1 THR A 12			-						
ATOM 125 HG1 THR A 12 -10.170 -3.766 30.096 ATOM 126 CG2 THR A 12 -11.579 -2.895 28.156 ATOM 127 N ILE A 13 -10.449 -0.932 25.594 ATOM 128 H ILE A 13 -10.409 -1.841 25.178 ATOM 129 CA ILE A 13 -10.409 -1.841 25.178 ATOM 130 C ILE A 13 -12.553 -0.292 24.693 ATOM 131 O ILE A 13 -12.553 -0.292 24.693 ATOM 132 CB ILE A 13 -12.935 -1.469 24.821 ATOM 133 CG1 ILE A 13 -10.432 0.364 23.511 ATOM 133 CG1 ILE A 13 -10.432 0.364 23.512 ATOM 134 CG2 ILE A 13 -8.986 0.806 23.747 ATOM 135 CD1 ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -13.209 1.622 24.481 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.088 0.877 22.719 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -18.826 1.342 26.810 ATOM 146 LHZ LYS A 14 -19.316 0.929 28.173 ATOM 146 LHZ LYS A 14 -19.316 0.929 28.743 ATOM 147 3HZ LYS A 14 -19.316 0.929 28.743 ATOM 146 LHZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 149 CD LYS A 14 -19.801 1.693 28.599 ATOM 140 CB LYS A 14 -19.801 1.693 28.599 ATOM 145 NZ LYS A 14 -19.801 1.693 28.599 ATOM 150 H ILE A 15 -15.535 0.005 28.743 ATOM 151 CA ILE A 15 -15.535 0.005 28.743 ATOM 152 CB ILE A 15 -15.642 0.347 20.040 ATOM 153 O ILE A 15 -15.642 0.347 20.041 ATOM 154 CB ILE A 15 -14.488 0.148 18.125 ATOM 155 CG ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 157 CD1 ILE A 15 -14.478 0.148 18.125 ATOM 159 H GLY A 16 -17.720 1.426 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.308 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 N GLY A 17 -19.908 0.204 21.334 ATOM 164 H GLY A 17 -19.908 0.204 21.334 ATOM 165 CA GLY A 16 -19.053 -0.143 18.745									
ATOM 126 CG2 THR A 12 -11.579 -2.895 28.156 ATOM 127 N ILE A 13 -10.449 -0.932 25.594 ATOM 128 H ILE A 13 -10.409 -1.841 25.178 ATOM 129 CA ILE A 13 -10.409 -1.841 25.178 ATOM 130 C ILE A 13 -12.553 -0.292 24.693 ATOM 131 O ILE A 13 -12.935 -1.469 24.821 ATOM 132 CB ILE A 13 -12.935 -1.469 24.821 ATOM 133 CG1 ILE A 13 -10.432 0.364 23.511 ATOM 134 CG2 ILE A 13 -8.986 0.806 23.747 ATOM 135 CD1 ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.470 0.658 24.438 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -14.838 0.330 24.100 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.088 0.877 22.719 ATOM 140 CB LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.826 1.342 26.810 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.316 0.929 28.173 ATOM 149 N ILE A 15 -15.806 0.670 28.743 ATOM 149 N ILE A 15 -15.806 0.670 28.743 ATOM 150 H ILE A 15 -15.806 0.150 22.078 ATOM 151 CA ILE A 15 -15.806 0.150 22.078 ATOM 152 C ILE A 15 -15.806 0.150 22.078 ATOM 153 O ILE A 15 -15.806 0.150 22.078 ATOM 154 CB ILE A 15 -15.806 0.150 22.082 ATOM 155 CG1 ILE A 15 -14.483 0.328 19.887 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 159 H GLY A 16 -17.720 1.426 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -19.897 -0.817 19.789 ATOM 160 CA GLY A 16 -19.053 -0.143 19.796 ATOM 164 H GLY A 16 -19.053 -0.143 19.799 ATOM 164 H GLY A 16 -19.053 -0.143 19.799 ATOM 164 H GLY A 16 -19.053 -0.143 19.799 ATOM 164 H GLY A 16 -19.053 -0.143 19.799 ATOM 164 H GLY A 16 -19.053 -0.143 19.308 ATOM 164 H GLY A 16 -19.084 -0.328 19.308 ATOM 164 H GLY A 16 -19.053 -0.143 19.799 ATOM 164 H GLY A 16 -19.053 -0.143 19.799 ATOM 164 H GLY A 16 -19.053 -0.143 19.799									
ATOM 127 N ILE A 13									
ATOM 128 H ILE A 13									
ATOM 129 CA ILE A 13 -11.112 0.133 24.882 ATOM 130 C ILE A 13 -12.553 -0.292 24.693 ATOM 131 O ILE A 13 -12.553 -1.469 24.821 ATOM 132 CB ILE A 13 -10.432 0.364 23.511 ATOM 133 CG1 ILE A 13 -10.466 -0.896 22.628 ATOM 134 CG2 ILE A 13 -8.986 0.806 23.747 ATOM 135 CD1 ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.470 0.658 24.438 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -14.838 0.330 24.100 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.088 0.877 22.779 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.078 0.146 26.166 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 148 2HZ LYS A 14 -19.801 1.693 28.599 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.642 0.347 20.400 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 157 CD1 ILE A 15 -14.478 0.148 18.125 ATOM 158 N GLY A 16 -17.720 1.426 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.308 ATOM 150 C GLY A 16 -17.720 1.426 19.308 ATOM 161 C GLY A 16 -17.720 1.426 19.308 ATOM 162 O GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 16 -19.997 -0.817 19.789 ATOM 164 H GLY A 16 -19.997 -0.817 19.789 ATOM 165 CA GLY A 16 -19.997 -0.817 19.789									
ATOM 130 C ILE A 13 -12.553 -0.292 24.693 ATOM 131 O ILE A 13 -12.935 -1.469 24.821 ATOM 132 CB ILE A 13 -10.432 0.364 23.511 ATOM 133 CG1 ILE A 13 -10.466 -0.896 22.628 ATOM 134 CG2 ILE A 13 -8.986 0.806 23.747 ATOM 135 CD1 ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -15.088 0.877 22.719 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -15.855 0.916 25.099 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.078 0.146 26.166 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.535 0.005 21.798 ATOM 151 CA ILE A 15 -15.806 0.916 22.078 ATOM 152 C ILE A 15 -15.806 0.916 22.078 ATOM 153 O ILE A 15 -15.806 0.916 22.078 ATOM 154 CB ILE A 15 -15.806 0.916 22.078 ATOM 155 CG1 ILE A 15 -15.806 0.916 22.078 ATOM 158 N GLY A 16 -17.15 -1.542 20.041 ATOM 158 N GLY A 16 -17.843 0.435 19.880 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 150 CG ILE A 15 -14.382 -0.132 19.880 ATOM 151 CA ILE A 15 -14.382 -0.132 19.639 ATOM 158 N GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 166 CA GLY A 16 -19.053 -0.143 18.745 ATOM 166 CA GLY A 16 -19.053 -0.0493 21.088 ATOM 161 C GLY A 16 -19.053 -0.0493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 16 -19.053 -0.0493 21.088									
ATOM 131 O LLE A 13 -12.935 -1.469 24.821 ATOM 132 CB ILE A 13 -10.432 0.364 23.511 ATOM 133 CG1 ILE A 13 -10.432 0.364 23.511 ATOM 134 CG2 ILE A 13 -10.466 -0.896 23.747 ATOM 135 CD1 ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.470 0.658 24.438 ATOM 137 H LYS A 14 -13.209 1.622 4.481 ATOM 138 CA LYS A 14 -13.209 1.622 23.75 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.078 0.146 26.166 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.642 0.347 20.400 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 153 O ILE A 15 -15.642 0.347 20.400 ATOM 155 CG1 ILE A 15 -14.478 0.148 1.25 ATOM 156 CG2 ILE A 15 -14.478 0.148 1.25 ATOM 157 CD1 ILE A 15 -14.478 0.148 1.25 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.308 ATOM 150 CG LY A 16 -19.957 -0.817 19.789 ATOM 161 C GLY A 16 -19.957 -0.817 19.789 ATOM 162 O GLY A 16 -19.957 -0.817 19.789 ATOM 163 N GLY A 16 -19.957 -0.817 19.789 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 1664 H GLY A 17 -19.038 0.204 21.334									
ATOM 132 CB ILE A 13 -10.432 0.364 23.511 ATOM 133 CG1 ILE A 13 -10.466 -0.896 22.628 ATOM 134 CG2 ILE A 13 -8.986 0.806 23.747 ATOM 135 CD1 ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.470 0.658 24.438 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -13.209 1.622 24.481 ATOM 139 C LYS A 14 -14.838 0.330 24.100 ATOM 140 O LYS A 14 -15.088 0.877 22.719 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -17.325 0.518 24.864 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 149 N ILE A 15 -15.642 0.347 20.400 ATOM 150 H ILE A 15 -15.642 0.347 20.400 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.082 -0.132 19.639 ATOM 157 CD1 ILE A 15 -14.082 -0.132 19.639 ATOM 158 N GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334									
ATOM 133 CG1 ILE A 13 -10.466 -0.896 22.628 ATOM 134 CG2 ILE A 13 -8.986 0.806 23.747 ATOM 135 CD1 ILE A 13 -9.755 -0.745 21.294 ATOM 136 N LYS A 14 -13.209 1.622 24.481 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -15.855 0.916 25.099 ATOM 143 CD LYS A 14 -17.325 0.518 24.864 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.642 0.347 20.400 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -15.642 0.347 20.400 ATOM 153 O ILE A 15 -15.642 0.347 20.400 ATOM 155 CG1 ILE A 15 -16.894 -0.328 19.887 ATOM 156 CG2 ILE A 15 -14.382 -0.132 19.693 ATOM 157 CD1 ILE A 15 -14.478 0.148 18.125 ATOM 158 N GLY A 16 -17.720 1.426 19.260 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334	MOTA								
ATOM 134 CG2 ILE A 13	ATOM								
ATOM 135 CD1 ILE A 13	MOTA			-					
ATOM 136 N LYS A 14 -13.470 0.658 24.438 ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -14.838 0.330 24.100 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.855 0.916 25.099 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -15.855 0.916 25.099 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.078 0.146 26.166 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.743 ATOM 148 2HZ LYS A 14 -19.535 0.057 22.078 ATOM 150 H ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.642 0.347 20.400 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 154 CB ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 157 CD1 ILE A 15 -14.382 -0.132 19.639 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -17.843 0.435 19.308 ATOM 161 C GLY A 16 -17.720 1.426 19.260 ATOM 163 N GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 16 -19.053 -0.143 18.745 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0	ATOM	134							
ATOM 137 H LYS A 14 -13.209 1.622 24.481 ATOM 138 CA LYS A 14 -14.838 0.330 24.100 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -14.859 2.059 22.375 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.316 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.670 28.743 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.806 -0.916 22.078 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.382 -0.132 19.639 ATOM 157 CD1 ILE A 15 -14.382 -0.132 19.639 ATOM 158 N GLY A 16 -17.720 1.426 19.260 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 150 CA GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.897 -0.817 19.789 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334	ATOM								
ATOM 138 CA LYS A 14 -14.838 0.330 24.100 ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -15.085 0.916 25.099 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.078 0.146 26.166 ATOM 145 NZ LYS A 14 -18.826 1.342 26.810 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 148 2HZ LYS A 14 -19.936 0.670 28.743 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.806 -0.916 22.078 ATOM 152 C ILE A 15 -15.642 0.347 20.400 ATOM 153 O ILE A 15 -15.642 0.347 20.400 ATOM 154 CB ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.382 -0.132 19.639 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 16 -19.053 -0.143 18.745 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334	ATOM	136	N						
ATOM 139 C LYS A 14 -15.088 0.877 22.719 ATOM 140 O LYS A 14 -14.859 2.059 22.375 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -15.855 0.916 25.099 ATOM 143 CD LYS A 14 -17.325 0.518 24.864 ATOM 144 CE LYS A 14 -18.078 0.146 26.166 ATOM 145 NZ LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 148 2HZ LYS A 14 -19.936 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.055 21.798 ATOM 150 H ILE A 15 -15.535 0.005 21.798 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -15.642 0.347 20.400 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.382 -0.132 19.639 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 160 CA GLY A 16 -17.720 1.426 19.260 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 16 -19.897 -0.817 19.789 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334	MOTA	137	H						
ATOM 140 O LYS A 14 -14.859 2.059 22.375 ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.8536 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.150 28.082 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.806 -0.916 22.078 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -17.115 -1.542 20.041 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 157 CD1 ILE A 15 -14.478 0.148 18.125 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.053 -0.817 19.789 ATOM 163 N GLY A 16 -19.053 -0.817 19.789 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160	MOTA	138	CA						
ATOM 141 CB LYS A 14 -15.855 0.916 25.099 ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 148 2HZ LYS A 14 -19.936 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.150 28.082 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -17.115 -1.542 20.041 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 157 CD1 ILE A 15 -14.082 -1.623 19.880 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 16 -19.053 -0.143 18.745 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334	ATOM	139	С						
ATOM 142 CG LYS A 14 -17.325 0.518 24.864 ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.936 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.150 28.082 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.642 0.347 20.400 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.082 -1.623 19.880 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334	MOTA	140	0						
ATOM 143 CD LYS A 14 -18.078 0.146 26.166 ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -18.536 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.150 28.082 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334	MOTA	141							
ATOM 144 CE LYS A 14 -18.826 1.342 26.810 ATOM 145 NZ LYS A 14 -19.316 0.929 28.173 ATOM 146 1HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -19.801 1.693 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.670 28.743 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.082 -1.623 19.880 ATOM 158 N GLY A 16 -17.720 1.426 19.260 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 16 -19.897 -0.817 19.789 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334	MOTA	142	CG		Α				
ATOM 145 NZ LYS A 14	MOTA	143							
ATOM 146 1HZ LYS A 14 -19.801 1.693 28.599 ATOM 147 3HZ LYS A 14 -18.536 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.150 28.082 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -17.115 -1.542 20.041 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 157 CD1 ILE A 15 -14.082 -1.623 19.880 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.843 0.435 19.308 ATOM 160 CA GLY A 16 -17.720 1.426 19.260 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 16 -19.897 -0.817 19.789 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334	ATOM	144	CE						
ATOM 147 3HZ LYS A 14 -18.536 0.670 28.743 ATOM 148 2HZ LYS A 14 -19.936 0.150 28.082 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 156 CG2 ILE A 15 -14.478 0.148 18.125 ATOM 157 CD1 ILE A 15 -14.082 -1.623 19.880 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -17.720 1.426 19.260 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 16 -20.774 -1.668 19.516 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334		145	NZ		Α				
ATOM 148 2HZ LYS A 14 -19.936 0.150 28.082 ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -17.115 -1.542 20.041 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160	ATOM	146	1HZ		Α				
ATOM 149 N ILE A 15 -15.535 0.005 21.798 ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -16.894 -0.328 19.887 ATOM 154 CB ILE A 15 -17.115 -1.542 20.041 ATOM 155 CG1 ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.082 -1.623 19.880 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.053 -0.143 18.745 ATOM 163 N GLY A 16 -20.774 -1.668 19.516 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160	MOTA	147	3HZ	LYS	Α	14			
ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -17.115 -1.542 20.041 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.082 -1.623 19.880 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.143 18.745 ATOM 162 O GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 16 -20.774 -1.668 19.516 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160	ATOM	148	2HZ	LYS	Α				
ATOM 150 H ILE A 15 -15.806 -0.916 22.078 ATOM 151 CA ILE A 15 -15.642 0.347 20.400 ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -17.115 -1.542 20.041 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.817 19.789 ATOM 162 O GLY A 16 -19.897 -0.817 19.789 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160	ATOM	149	N						
ATOM 151 CA 1LE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -17.115 -1.542 20.041 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.053 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.038 0.204 21.334 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160	ATOM	150	H	ILE	Α	15			
ATOM 152 C ILE A 15 -16.894 -0.328 19.887 ATOM 153 O ILE A 15 -17.115 -1.542 20.041 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160	MOTA	151	CA	ILE	Α				
ATOM 153 O ILE A 15 -17.115 -1.542 20.041 ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160		152	С	ILE	Α	15			
ATOM 154 CB ILE A 15 -14.382 -0.132 19.639 ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160			0	ILE	Α	15			
ATOM 155 CG1 ILE A 15 -14.478 0.148 18.125 ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160		154	CB	ΙĽΕ	Α	15			
ATOM 156 CG2 ILE A 15 -14.082 -1.623 19.880 ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160			CG1	ILE	Α	15			
ATOM 157 CD1 ILE A 15 -14.237 1.603 17.796 ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160			CG2	ILE	Α	15			
ATOM 158 N GLY A 16 -17.843 0.435 19.308 ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160			CD1	ILE	Α	15			
ATOM 159 H GLY A 16 -17.720 1.426 19.260 ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160		158	N	GLY	A	16			
ATOM 160 CA GLY A 16 -19.053 -0.143 18.745 ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160				GLY	Α	16			
ATOM 161 C GLY A 16 -19.897 -0.817 19.789 ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160						16			
ATOM 162 O GLY A 16 -20.774 -1.668 19.516 ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160						16	-19.897		
ATOM 163 N GLY A 17 -19.712 -0.493 21.088 ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160						16			
ATOM 164 H GLY A 17 -19.038 0.204 21.334 ATOM 165 CA GLY A 17 -20.464 -1.126 22.160									
ATOM 165 CA GLY A 17 -20.464 -1.126 22.160									
A1011 200 GT 10 710 7 725 77 653									
ATOM 100 C GDI K I'		166	Ċ.			17	-19.718	-2.335	22.653
ATOM 167 O GLY A 17 -20.147 -3.098 23.540						17	-20.147	-3.098	23.540

FIG. I IC

WO 01/35316 PCT/US00/30863

17/46 22.121 -2.591 -18.507 18 GLN A 168 N MOTA -1.900 21.554 -18.059 GLN A 18 Η 169 MOTA 22.326 -3.830 -17.806 GLN A 18 **ATOM** 170 CA -3.549 23.123 -16.552 18 C GLN A 171 **ATOM** -2.508 22.945 -15.887 GLN A 18 172 0 MOTA 20.928 -4.294 -17.393 GLN A 18 MOTA 173 CB -5.734 20.788 -16.911 GLN A 18 174 CG MOTA 20.613 -6.728 -18.018 GLN A 18 CD 175 **ATOM** -6.574 21.152 OE1 GLN A 18 -19.131 176 MOTA -7.773 19.857 -17.722 NE2 GLN A 18 177 MOTA -8.484 19.689 -18.404 178 1HE2 GLN A 18 MOTA -7.860 19.448 -16.814 2HE2 GLN A 18 179 MOTA -4.397 24.087 -16.133 LEU A 19 180 N MOTA 24.312 -5.202 LEU A 19 -16.682 181 Н MOTA 24.808 -4.178 -14.909 LEU A 19 CA 182 MOTA -4.912 24.090 -13.799 LEU A 19 183 С ATOM 23.558 -6.018 -13.989 LEU A 19 0 MOTA 184 -4.714 26.254 -14.982 LEU A 19 185 CB MOTA -3.778 -15.490 27.374 LEU A 19 CG 186 MOTA 26.856 -16.392 -2.639 CD1 LEU A 19 187 MOTA -4.516 28.465 -16.208 CD2 LEU A 19 MOTA 188 23.978 -4.372 -12.603 LYS A 20 189 N MOTA 24.324 -3.448 -12.442 LYS A 20 190 Н MOTA -5.082 23.365 -11.507 LYS A 20 CA MOTA 191 24.062 -4.618 -10.266 LYS A 20 **MOTA** 192 C -3.611 24.816 -10.228 20 LYS A MOTA 193 0 -4.798 21.875 -11.397 20 LYS A 194 CB MOTA 21.100 -5.356 -12.558 20 CG LYS A 195 ATOM 19.615 -4.988 -12.537 LYS A 20 CD **ATOM** 196 18.827 -5.958 -13.414 LYS A 20 197 CE MOTA 18.639 -7.208 -12.681 NZ LYS A 20 198 MOTA 18.123 -13.247 -7.852 199 1HZ LYS A 20 MOTA -7.601 19.531 -12.458 200 3HZ LYS A 20 ATOM -7.027 18.134 -11.837 LYS A 20 MOTA 201 2HZ -5.357 23.893 -9.150 202 GLU A 21 ATOM N 23.338 -6.188 -9.185 GLU A 21 203 Н ATOM 24.486 -4.997 -7.890 GLU A 21 204 CA MOTA 23.390 -4.462 -7.001 GLU A 21 MOTA 205 C 22.258 -4.992 -6.970 0 GLU A 21 MOTA 206 25.051 -6.260 -7.268 207 CB GLU A 21 **ATOM** 25.480 -6.140 -5.835 CG GLU A 21 208 MOTA -7.352 26.275 -5.405 GLU A 21 **ATOM** 209 CD 27.508 -7.343 -5.624 210 OE1 GLU A 21 MOTA -8.309 25.684 -4.852 OE2 GLU A 211 21 MOTA 23.595 -3.369 -6.239 22 212 N ALA A MOTA -2.938 24.497 -6.223 ALA A 22 MOTA 213 Н -2.781 22.520 -5.419 ALA A 22 214 CA MOTA -2.255 23.114 -4.138 22 C ALA A **ATOM** 215 -1.914 24.314 -3.985 22 0 ALA A MOTA 216 -6.134 -1.657 21.821 ALA A 22 CB ATOM 217 -3.121 -2.091 22.240 LEU A 23 218 atom N

FIG. ID SUBSTITUTE SHEET (RULE 26)

23

23

23

23

23

LEU A

LEU A

LEU A

LEU A

LEU A

219

220

221

222

223

ATOM

MOTA

"ATOM

MOTA

MOTA

Н

C

0

CB

CA

-3.279

-1.797

-1.660

-2.020

-0.814

-2.236

-1.712

-0.230

-2.486

0.349

21.263

22.640

22.443

21.402

21.732

					18/46			
					-	0 705	-2.448	21.991
ATOM	224	CG	LEU	A	23	0.705 1.088	-3.400	23.124
ATOM	225	CD1	LEU LEU	A	23 23	1.462	-2.878	20.708
ATOM	226	CD2	LEU		24	-1.192	0.530	23.463
ATOM	227	N	LEU		24	-1.015	0.110	24.353
ATOM	228	H CA	LEU	Δ	24	-0.935	1.952	23.305
ATOM	229 230	CA	LEU		24	0.403	2.089	
ATOM ATOM	231	0	LEU		24	1.471	1.717	23.130
ATOM	232	СВ	LEU		24	-0.921	2.609	24.681
ATOM ATOM	233	CG	LEU		24	-2.220	2.492	25.477
ATOM	234	CD1	LEU		24	-2.063	3.291	26.772
ATOM	235	CD2	LEU		24	-3.419	3.000	24.691
ATOM	236	N	ASP	Α	25	0.454	2.590	21.397
ATOM	237	H	ASP	Α	25	-0.334	3.085	21.032
ATOM	238	CA	ASP	Α	25	1.642	2.423	20.605
ATOM	239	C	ASP		25	2.130	3.750	20.059
ATOM	240	0	ASP		25	1.568	4.320	19.110
ATOM	241	CB	ASP		25	1.263	1.435	19.486
ATOM	242	CG	ASP		25	2.428	1.051	18.561 18.729
ATOM	243	OD1		A	25	3.546	1.540 0.241	17.658
ATOM	244	OD2	ASP		25	2.164	4.337	20.605
ATOM	245	N	THR		26	3.203 3.694	3.880	21.346
ATOM	246	H	THR		26 26	3.691	5.652	20.144
ATOM	247	CA	THR THR		26	4.397	5.583	18.778
ATOM	248	0	THR		26	4.642	6.587	18.079
ATOM	249 250	CB	THR		26	4.596	6.219	21.217
ATOM ATOM	250 251	OG1	THR		26	5.716	5.324	21.386
ATOM	252	HG1	THR		26	6.332	5.676	22.091
ATOM	253	CG2	THR		26	3.878	6.320	22.577
ATOM	254	N	GLY		27	4.757	4.377	18.298
ATOM	255	Н	GLY	Α	27	4.526	3.550	18.811
ATOM	256	CA	GLY	A	27	5.481	4.233	17.040
ATOM	257	C	GLY		27	4.520	4.190	15.886
MOTA	258	0	GLY		27	4.908	4.242	14.696
MOTA	259	N	ALA		28	3.197	4.084	16.117
MOTA	260	H	ALA		28	2.856	4.091	17.057
MOTA	261	CA	ALA		28	2.213	3.955	15.018 14.750
ATOM	262	C	ALA		28	1.598	5.299 5.982	15.650
ATOM	263	0	ALA		28	1.062 1.117	2.980	15.390
ATOM	264	CB	ALA ASP		28 29	1.503	5.744	13.490
ATOM	265 266	N H	ASP		29	1.912	5.216	12.746
ATOM ATOM	267	CA	ASP		29	0.810	6.984	13.213
ATOM	268	C	ASP		29	-0.666	6.724	13.327
ATOM	269	ŏ	ASP			-1.488	7.637	13.568
ATOM	270	CB	ASP		29	1.009	7.433	11.775
ATOM	271	CG	ASP		29	2.439	7.882	11.412
ATOM	272	OD1	ASP		29	3.360	7.856	12.269
MOTA	273	OD2	asp	A	29	2.606	8.253	10.252
ATOM	274	N	ASP		30	-1.143	5.517	12.990
ATOM	275	H	ASP		30	-0.508	4.769	12.800
ATOM	276	CA	ASP		30	-2.579	5.245	12.887
ATOM	277	C	ASP		30	-3.057	4.208	13.867 14.546
ATOM	278	0	ASP		30	-2.284	3.483	14.546
MOTA	279	CB	ASP	A	30	-2.896	4.758	11.730

1	9	1	L.	6
		,	-	v

ATOM	280	CG	ASP	Α	30	-2.495	5.768	10.425
MOTA	281	OD1	ASP	Α	30	-3.067	6.871	10.423
ATOM	282	OD2	ASP	Α	30	-1.596	5.494	9.618
ATOM	283	N	THR	Α	31	-4.393	4.076	14.002
ATOM	284	H	THR		31	-5.004	4.700	13.515
MOTA	285	CA	THR		31	-5.059	3.062	14.829
ATOM	286	C	THR		31	-5.565	1.967	
ATOM	287	0	THR		31	-6.223	2.169	12.870
MOTA	288	CB	THR		31	-6.212	3.725	15.566
ATOM	289	OG1			31	-5.668	4.667	16.474
ATOM	290	HG1	THR		31	-6.403	5.122	16.976
ATOM	291	CG2	THR		31	-7.044	2.702	16.389
ATOM	292	N	VAL		32	-5.187	0.713	14.235
ATOM	293	H		Α	32	-4.649	0.555	15.063
ATOM	294	CA		A	32	-5.517	-0.462	13.437
ATOM	295	Ċ.		A	32	-6.092	-1.506	14.365
MOTA	296	0	VAL		32	-5.502	-1.957	15.365
MOTA	297	CB	VAL		32	-4.260	-1.064	12.757
ATOM	298	CG1	VAL		32	-4.667	-2.136	11.735
MOTA	299	CG2	VAL		32	-3.422	0.017	12.032
ATOM	300	N	LEU		33	-7.352	-1.923	14.119 13.361
MOTA	301	H	LEU		33	-7.867 -7.982	-1.523 -2.940	14.929
ATOM	302	CA	LEU		33	-7.982 -8.174	-4.203	14.323
ATOM	303	С		A	33	-8.268	-4.203	12.853
MOTA	304	0	LEU		33	-9.336	-2.477	15.408
ATOM	305	CB		A	33 33	-9.292	-1.149	16.127
ATOM	306	CG CD1		A A	33	-10.710	-0.747	16.485
ATOM	307	CD1	LEU		33	-8.348	-1.139	17.347
ATOM ATOM	308 309	N	GLU		34	-8.296	-5.319	14.782
ATOM	310	H ·	GLU		34	-8.244	-5.302	15.780
ATOM	311	CA	GLU		34	-8.503	-6.551	14.086
ATOM	312	C	GLU		34	-9.909	-6.549	13.510
ATOM	313	ō	GLU		34	-10.808	-5.717	13.795
ATOM	314	СB	GLU		34	-8.265	-7.750	15.010
ATOM	315	CG	GLU		34	-9.259	-7.791	16.165
ATOM	316	CD	GLU		34	-8.763	-8.552	17.404
ATOM	317	OE1	GLU		34	-7.670	-9.193	17.368
MOTA	318	OE2	GLU		34	-9.482	-8.497	18.407
ATOM	319	N	GLU	Α	35	-10.152	-7.480	12.568
MOTA	320	H	GLU		35	-9.485	-8.208	12.407
ATOM	321	CA	GLU	Α	35	-11.352	-7.466	11.773
ATOM	322	C	GLU	Α	35	-12.631	-7.520	12.571
MOTA	323	0	GLU	Α	35	-12.814	-8.294	13.528
MOTA	324	CB	GLU	Α	35	-11.237	-8.536	10.707
ATOM	325	CG	GLU		35	-9.945	-8.280	9.907
ATOM	326	CD	GLU		35	-9.872	-8.872	8.486
MOTA	327	OE1	GLU		35	-10.612	-8.401	7.603
ATOM	328	OE2	GLU		35	-9.024	-9.776	8.261
ATOM	329	N		A	36		-6.598	12.278
atom	330	H		A	36		-5.967	11.515
ATOM	331	CA		A	36		-6.495	13.052
ATOM	332	C		A	36		-5.635	12.271
MOTA	333	0		Α	36		-4.828	11.371
ATOM	334	CB		A	36		-5.845	14.428
ATOM	335	CG	MET	A	36	-14.279	-4.353	14.417

FIG. 1 IF

SUBSTITUTE SHEET (RULE 26)

				20/4	6		
ATOM	336	SD	MET A		-14.251	-3.718	16.099
ATOM	337	CE	MET A	36	-12.487	-3.846	16.409
ATOM	338	N	SER A	37	-17.130	-5.776	12.590
ATOM	339	H	SER A	37	-17.399	-6.431	13.296
ATOM	340	CA	SER A		-18.155	-5.005	11.940
ATOM	341	C	SER A		-18.286	-3.693	12.657
ATOM	342	0	SER A		-18.593	-3.624	. 13.865
ATOM	343	CB	SER A		-19.506	-5.688	12.032
ATOM	344	OG	SER A		-19.455	-7.054	11.715
ATOM	345	HG	SER A		-20.367	-7.457	11.791
ATOM	346	N	LEU A		-18.185	-2.569	11.933
ATOM	347	H	LEU A		-17.956	-2.625	10.952
ATOM	348	CA	LEU A		-18.557	-1.247	12.465
ATOM	349	C	LEU A		-19.630	-0.605	11.572
ATOM	35:0	ō	LEU A		-19.706	-0.939	10.391
ATOM	351	CB	LEU A		-17.315	-0.346	12.588
ATOM	352	CG	LEU A		-16.246	-0.818	13.596
ATOM	353	CD1	LEU A		-14.998	0.073	13.489
ATŌM	354	CD2	LEU A		-16.756	-0.787	15.046
ATOM	355	N	PRO A		-20.455	0.321	12.108
ATOM	356	CA	PRO A		-21.460	1.053	11.339
ATOM	357	C	PRO A		-20.824	2.176	10.502
ATOM	358	ō	PRO A		-19.654	2.519	10.685
ATOM	359	СВ	PRO A		-22.430	1.607	12.389
ATOM	360	CG	PRO A		-21.531	1.845	13.600
ATOM	361	CD	PRO A	39	-20.539	0.686	13.517
ATOM	362	N	GLY A	40	-21.620	2.749	9.586
ATOM	363	H	GLY A	40	-22.569	2.417	9.493
ATOM	364	CA	GLY A	40	-21.203	3.811	8.678
ATOM	365	C	GLY A	40	-20.836	3.262	7.298
ATOM	366	ō	GLY A	40	-21.405		6.845
ATOM	367	N	LYS A	41	-19.895	3.945	6.631
ATOM	368	H	LYS A	41	-19.496	4.761	7.071
ATOM	369	CA	LYS A	41	-19.323	3.558	5.343
ATOM	370	C	LYS A	41	-17.798	3.757	5.371
ATOM	371	0	LYS A	41	-17.263	4.462	6.229
ATOM	372	CB	LYS A	41	-20.025	4.352	4.224
ATOM	373	CG	LYS A	41	-19.703	3.839	2.810
ATOM	374	CD	LYS A	41	-20.610	4.486	1.757
ATOM	375	CE	LYS A	41	-20.240	3.964	0.366
ATOM	376	NZ	LYS A	41	-21.097	4.552	-0.678
ATOM	377	1HZ	LYS A	41	-20.824	4.189	-1.580
ATOM	378	3HZ	LYS A	41	-20.993	5.556	-0.673
ATOM	379	2HZ	LYS A	41	-22.061	4.311	-0.498
ATOM	380	N	TRP A	42	-17.104	3.091	4.439
ATOM	381	H	TRP A	42	-17.620	2.548	3.762
ATOM	382	CA	TRP A	42	-15.654	2.932	4.423
ATOM	383	С	TRP A	42	-15.105	2.852	2.994
ATOM	384	0	TRP A	42	-15.845	2.702	2.021
ATOM	385	CB	TRP A	42	-15.279	1.675	5.236
atom	386	CG	TRP A	42	-16.214	0.514	5.094
ATOM	387	CD1	TRP A	42	-16.230	-0.402	4.101
ATOM	388	CD2	TRP A	42	-17.355	0.203	5.942
ATOM	389	NEI	TRP A	42	-17.297	-1.260	4.281
ATOM	390	HE1	TRP A	42	-17.504	-2.015	3.644
ATOM	391	CE2	TRP A	42	-18.045	-0.914	5.389

FIG. 1 IG SUBSTITUTE SHEET (RULE 26)

2	1	•	,	
Z	Į	/	4	O

MOTA	392	CE3	TRP A	42	-17.896	0.792	7.103
MOTA	393	CZ2		42	-19.224	-1.421	5.959
ATOM	394	CZ3	TRP A	42	-19.077	0.298	7.675
ATOM	395	CH2	TRP A	42	-19.741	-0.806	7.112
MOTA	396	N	LYS A	43	-13.771	2.932	2.911
ATOM	397	H	LYS A	43	-13.260	3.058	3.773
ATOM	398	CA	LYS A	43	-12.951	2.802-	
MOTA	399	С	LYS A	43	-11.773	1.859	2.012
ATOM	400	0	LYS A	43	-11.359	1.760	3.166
ATOM	401	CB	LYS A	43	-12.451	4.193	1.270
MOTA	402	CG	LYS A	43	-11.724	4.979	2.383
ATOM	403	CD	LYS A	43	-11.060	6.267	1.873
ATOM	404	CE	LYS A	43	-9.784	6.001	1.065
ATOM	405	NZ	LYS A	43	-8.700	5.458	1.903
MOTA	4 0.6	1HZ	LYS A	43	-7.876	5.315	1.338
ATOM	407	3HZ	LYS A	43	-8.993	4.576	2.300
ATOM	408	2HZ	LYS A	43	-8.493	6.108	2.647
MOTA	409	N	PRO A	44	-11.177	1.197	1.004
MOTA	410	CA	PRO A	44	-9.947	0.435	1.187
MOTA	411	C	PRO A	44	-8.760	1.392	1.379
MOTA	412	0	PRO A	44	-8.711	2.434	0.720
ATOM	413	CB	PRO A	44	-9.808	-0.393	-0.095
ATOM	414	CG	PRO A	44	-10.501	0.458	-1.159
ATOM	415	CD	PRO A	44	-11.630 -7.790	1.132 1.030	-0.380 2.240
ATOM	416	N	LYS A	45 45	-7.790	0.227	2.824
ATOM	417 418	H CA	LYS A LYS A	45 45	-7.912 -6.547	1.747	2.314
ATOM ATOM	419	CA	LYS A	45	-5.493	0.683	2.507
ATOM	420	0	LYS A	45	-5.780	-0.470	2.869
ATOM	421	CB	LYS A	45	-6.594	2.699	3.524
ATOM	4.22	CG	LYS A	45	-5.463	3.744	3.609
ATOM	423	CD	LYS A	45	-5.340	4.289	5.052
ATOM	424	CE	LYS A	45	-4.262	5.383	5.204
ATOM	425	NZ	LYS A	45	-2.907	4.911	4.916
ATOM	426	1HZ	LYS A	45	-2.260	5.664	5.032
ATOM	427	3HZ	LYS A	45	-2.864	4.577	3.975
ATOM	428	2HZ	LYS A	45	-2.672	4.169	5.544
ATOM	429	N	MET A	46	-4.224	0.949	2.193
ATOM	430	Н	MET A	46	-3.998	1.805	1.728
ATOM	431	CA	MET A	46	-3.157	0.027	2.509
ATOM	432	С	MET A	46	-2.417	0.701	3.627
MOTA	433	0	MET A	46	-2.259	1.937	3.634
MOTA	434	CB	MET A	46	-2.166	-0.088	1.379
MOTA	435	CG	MET A	46	-2.782	-0.366	0.053
MOTA	436	SD	MET A	46	-3.076	-2.108	-0.118
MOTA	437	CE	MET A	46	-1.417	-2.652	-0.186
MOTA	438	N	ILE A	47	-1.827	-0.016	4.586
MOTA	439	H	ILE A	47	-2.010	-0.997	4.655
ATOM	440	CA	ILE A	47	-0.922	0.586	5.539
ATOM	441	C	ILE A	47	0.233	-0.372	5.654
MOTA	442	0	ILE A	47	0.135	-1.584	5.356
ATOM	443	CB	ILE A	47	-1.550	0.836	6.923
ATOM	444	CG1	ILE A	47	-2.459	-0.301	7.354
ATOM	445	CG2	ILE A	47	-2.248	2.164	6.995
ATOM	446	CD1	ILE A	47	-1.724	-1.336	8.111
MOTA	447	N	GLY A	48	1.420	0.089	6.043

FIG. 1 IH SUBSTITUTE SHEET (RULE 26)

\mathbf{a}	1	,	10
/	/	/	40

			•		•			
ATOM	448	Н	GLY A	48]	L.50 ⁵ 9	1.040	6.339
ATOM	449	CA	GLY A	48	2	2.584	-0.753	6.048
ATOM	450	C	GLY A	48		3.280		7.376
ATOM	451	0	GLY A	48		.050		8.265
MOTA	452	N	GLY A	49	4	.197	-1.617	7.603
ATOM	453	H	GLY A		4	.375	-2.308	6.902
MOTA	454	CA	GLY A		4	.936	-1.684	
ATOM	455	C	GLY A			.105		8.533
ATOM	456	0	GLY A			.482		7.370
ATOM	457	N	ILE A			.761		9.552
ATOM	458	H	ILE A			.552	-2.908	10.493
ATOM	459	CA	ILE A			.772	-4.184	9.344
ATOM	460	C	ILE A			.148	-5.317	8.566
ATOM	461	0	ILE A			.981	-5.734	8.772
ATOM	462	CB	ILE A			.258	-4.686	10.722
MOTA	463	CG1				.257	-3.714	11.382
MOTA	464	CG2				.813	-6.134	10.693
ATOM	465	CDI				.580	-3.498	10.628
ATOM	466	N	GLY A	51		.847	-5.891	7.596
ATOM	467	H	GLY A	51		.772	-5.569	7.395
ATOM	468	CA	GLY A	51		. 265	-6.966	6.850
ATOM	469	C	GLY A	51		.519	-6.559	5.591
ATOM	470	0	GLY A	51		.430	-7.318	4.634
ATOM ATOM	471	N	GLY A	52 53		. 886	-5.375	5.517
ATOM	472	H CA	GLY A	52 52		. 990	-4.710	6.257
ATOM	473 474	CA	GLY A GLY A	52 52		108	-5.227	4.320
ATOM	475	0	GLY A	52 52		832 654	-4.415	4.516 5.467
ATOM	476	N	PHE A	52 53		886	-3.624	3.559
ATOM	477	H	PHE A	53		013	-4.518 -5.161	2.804
ATOM	478	CA	PHE A	53		653	-3.720	3.566
ATOM	479	C	PHE A	53		494	-4.651	3.783
ATOM	480	Õ	PHE A	53		448	-5.816	3.336
ATOM	481	CB	PHE A	53		424	-3.022	2.221
ATOM	482	CG	PHE A	53		363	-1.896	2.008
ATOM	483	CD1	PHE A	5 3		615	-2.135	1.447
ATOM	484	CD2	PHE A	53		011	-0.608	2.414
ATOM	485	CE1	PHE A	53		514	-1.087	1.275
ATOM	486	CE2	PHE A	53		925	0.446	2.237
ATOM	487	CZ	PHE A	53		172	0.202	1.668
ATOM	488	N	ILE A	54		554	-4.173	4.439
ATOM	489	H	ILE A	54	-0.		-3.285	4.895
ATOM	490	CA	ILE A	54	-1.	789	-4.911	4.509
ATOM	491	С	ILE A	54	-2.	903	-3.995	4.033
ATOM	492	0	ILE A	54	-2.		-2.770	3.855
ATOM	493	CB	ILE A	54	-2.	034	-5.535	5.904
ATOM	494	CG1	ILE A	54	-2.	343	-4.481	6.988
ATOM	495	CG2	ILE A	54	-0.	799	-6.318	6.314
ATOM	496	CD1	ILE A	54	-3.0		-5.089	8.246
ATOM	497	N	LYS A	55	-4.0	029	-4.577	3.560
ATOM	498	H	LYS A	55	-4.(-5.574	3.501
ATOM	499	CA	Lys a	55	-5.1		-3.798	3.129
ATOM	500	C	LYS A	55	-6.1		-3.726	4.300
ATOM	501	0	LYS A	55	-6.4		-4.707	5.023
ATOM	502	CB	LYS A	55	-5.9		-4.461	1.938
ATOM	503	CG	LYS A	55	-6.8	353	-3.547	1.106

FIG. I II

SUBSTITUTE SHEET (RULE 26)

ATOM	504	CD	LYS .	A 55	-8.267	-3.332	1.714
ATOM	505	CE	LYS	A 55	-9.303	-4.392	1.301
ATOM	506	NZ	LYS .	A 55	-10.521	-4.453	2.192
ATOM	507	1HZ	LYS 2	A 55	-11.142	-5.162	1.859
ATOM	508	3HZ		A 55	-10.987	-3.569	2.180
ATOM	509			A 55	-10.240	-4.669	3.127
ATOM	510		VAL A	A 56	-6.599	-2.509	4.619
ATOM	511	Н	VAL A		-6.337	-1.713	4.073
ATOM	512	CA	VAL A		-7.494	-2.311	5.735
ATOM	513	С	VAL A		-8.711	-1.584	5.236
ATOM	514	0	VAL A		-8.767	-1.029	4.114
ATOM	515	CB	VAL A	•	-6.759	-1.475	6.812
ATOM	516	CG1			-5.569	-2.209	7.385
ATOM	517	CG2			-6.287	-0.108	6.268
MOTA	518	N	ARG A		-9.784	-1.539	6.005
ATOM	519	H	ARG A		-9.835	-2.117	6.819
ATOM	520	CA	ARG A		-10.855	-0.648	5.638
ATOM	521	C		57	-10.738	0.534	6.554
ATOM	522	ō	ARG A		-10.558	0.449	7.789
ATOM	523	CB	ARG A		-12.219	-1.271	5.835
ATOM	524	CG	ARG A		-12.480	-2.452	4.952
ATOM	525	CD	ARG A		-13.834	-3.051	5.195
ATOM	526	NE	ARG A		-14.122	-4.137	4.270
ATOM	527	HE	ARG A		-13.442	-4.347	3.568
ATOM	528	CZ	ARG A		-15.243	-4.851	4.324
ATOM	529	NH1	ARG A		-16.175	-4.624	5.243
ATOM	530	2HH1	ARG A		-16.044	-3.899	5.920
ATOM	531	1HH1	ARG A		-17.008	-5.178	5.258
ATOM	532	NH2	ARG A		-15.433	-5.822	3.434
ATOM	533	1HH2	ARG A		-16.270	-6.368	3.461
ATOM	534	2HH2	ARG A		-14.738	-6.006	2.738
ATOM	535	N	GLN A		-10.881	1.741	6.036
ATOM	536	H	GLN A		-11.030	1.844	5.053
ATOM	537	CA	GLN A		-10.830	2.922	6.839
ATOM	538	C	GLN A		-12.231	3.342	7.205
ATOM	539	ŏ	GLN A		-13.106	3.608	6.359
ATOM	540	CB	GLN A		-10.208	4.038	6.030
ATOM	541	CG	GLN A		-10.055	5.293	6.817
ATOM	542	CD	GLN A		-9.632	6.411	5.927
ATOM	543	OE1	GLN A	58	-10.379	7.334	5.662
ATOM	544	NE2	GLN A		-8.412	6.303	5.437
ATOM	545	1HE2	GLN A	58	-8.047	7.009	4.830
ATOM	546	2HE2	GLN A	58	-7.843	5.514	5.668
ATOM	547	N	TYR A	59	-12.527	3.516	8.509
ATOM	548	Н	TYR A	59	-11.877	3.219	9.209
ATOM	549	CA	TYR A	59	-13.769	4.125	8.933
ATOM	550	C	TYR A	59	-13.411	5.452	9.565
ATOM	551	Ö	TYR A	59	-12.416	5.592	10.310
ATOM	552	СB	TYR A	59	-14.517	3.252	9.957
ATOM	553	CG	TYR A	59	-14.287	1.770	9.723
ATOM	554	CD1	TYR A	59	-13.007	1.269	9.457
ATOM	555	CD2	TYR A	59	-15.346	0.865	9.766
ATOM	556	CE1	TYR A	59	-12.797	-0.092	9.240
ATOM	557	CE2	TYR A	59	-15.148	-0.494	9.551
ATOM	558	CZ	TYR A	59	-13.873	-0.972	9.287
ATOM	559	ОН	TYR A	59	-13.721	-2.311	9.079

FIG. 1 J SUBSTITUTE SHEET (RULE 26)

				2	4/46			
MOTA	560	нн	TYR	Α	59	-14.606	-2.771	9.154
ATOM	561	N	ASP	Α	60	-14.151	6.542	9.300
ATOM	562	H	ASP	Α	60	-14.954	6.464	8.709
ATOM	563	CA	ASP	Α	60	-13.822	7.836	9.846
ATOM	564	C	ASP	Α	60	-14.782	8.226	10.947
ATOM	565	Ö	ASP		60	-15.941	7.765	11.053
ATOM	566	СB			60	-13.861	8.942-	8.769
ATOM	567	CG	ASP	Α	60	-12.735	8.830	7.725
ATOM	568	OD1		Α	60	-11.545	8.874	8.075
ATOM	569	OD2		A	60	-13.060	8.702	
ATOM	570	N	GLN		61	-14.339	9.154	11.833
MOTA	571	Н	GLN		61	-13.385	9.451	11.804
ATOM	572	CA	GLN		61	-15.151	9.804	12.885
ATOM	573	C	GLN		61	-15.839	8.803	13.802
ATOM	574	0	GLN		61	-17.008	8.893	14.229
ATOM	575	CB		A	61	-16.097	10.908	12.338
ATOM	576	CG		A	61	-16.239	12.133	13.262
	577	CD	GLN		61	-16.910	13.366	12.629
ATOM		OE1	GLN		61	-16.509	13.854	11.586
ATOM	578		GLN		61	-17.937	13.887	13.292
ATOM	579	NE2	GLN		61	-18.416	14.689	12.934
ATOM	580	1HE2			61	-18.239	13.482	14.155
ATOM	581	2HE2	-	A		-15.060	7.760	14.175
ATOM	582	N		A	62		7.760	13.862
MOTA	583	H	ILE		62	-14.111 -15.557	6.705	15.015
ATOM	584	CA		A	62			16.447
ATOM	585	С		A	62	-15.251	7.057 7.613	16.837
ATOM	586	0		A	62	-14.198		14.653
ATOM	587	CB		A	62	-14.829	5.397	13.258
ATOM	588	CG1		A	62	-15.253	4.966	
ATOM	589	CG2		A	62	-15.106	4.271	15.675 13.116
ATOM	590	CD1		A	62	-16.779	4.788	17.320
ATOM	591	N	LEU		63	-16.242	6.807	17.000
ATOM	592	H	LEU		63 .	-17.089	6.383 7.131	18.719
ATOM	593	CA		A	63	-16.127		19.425
ATOM	594	C		A	63	-15.518	5.942	
ATOM	595	0		A	63	-15.869	4.753	19.269
ATOM	596	CB	LEU		63	-17.512	7.428	19.282
ATOM	597	CG	LEU		63	-17.660	7.598	20.813
ATOM	598	CD1	LEU		63	-16.711	8.632	21.404 21.201
ATOM	599	CD2	LEU		63	-19.089	7.963	20.219
ATOM	600	N		Α	64	-14.511	6.211	
ATOM	601	H	ILE		64	-14.185	7.153	20.305
ATOM	602	CA		A	64	-13.862	5.178	20.972
ATOM	603	C		A	64	-13.529	5.744	22.325
ATOM	604	0_	ILE		64	-13.396	6.959	22.602
ATOM	605	CB		Α	64	-12.618	4.716	20.231
ATOM	606	CG1	ILE		64	-11.925	3.573	20.949
MOTA	607	CG2	ILE		64	-11.690	5.865	19.950
ATOM	608	CD1	ILE		64	-10.905	2.888	20.062
ATOM	609	Ŋ	GLU		65	-13.396	4.815	23.294
atom	610	H	GLU		65	-13.443	3.844	23.059
ATOM	611	CA	GLU		65	-13.186	5.174	24.670
ATOM	612	C	GLU		65	-12.024	4.360	25.165
ATOM	613	0_	GLU		65	-11.943	3.112	25.056
ATOM	614	CB	GLU		65	-14.459	4.823	25.405
ATOM	615	CG	GLU	A	65	-14.739	5.610	26.646

FIG. 1 K SUBSTITUTE SHEET (RULE 26)

				2	5/46			
MOTA	616	CD	GLU	Α	65	-16.131	5.353	27.115
ATOM	617	OE1	GLU	Α	65	-17.090	5.785	26.413
ATOM	618	OE2	GLU	Α	65	-16.269	4.708	28.163
MOTA	619	N	ILE	Α	66	-10.971	5.008	25.610
ATOM	620	H	ILE	Α	66	-11.009	6.002	25.717
ATOM	621	CA	ILE	Α	66	-9.762	4.317	25.947
ATOM	622	С	ILE	Α	66	-9.571	4.586-	
ATOM	623	0	ILE	Α	66	-9.422	5.732	27.880
ATOM	624	CB	ILE	A	66	-8.600	4.907	25.126
MOTA	625	CG1	ILE	Α	66	-8.838	4.669	23.633
MOTA	626	CG2	ILE	Α	66	-7.231	4.326	25.554
ATOM	627	CD1	ILE	A	66	-8.951	5.982	22.856
ATOM	628	N	CYS	Α	67	-9.776	3.567	28.261
MOTA	629	H	CYS	A	67	-9.989	2.659	27.902
ATOM	63.0	CA	CYS	A	67	-9.698	3.740	29.687
ATOM	631	C		A	67	-10.673	4.871	30.088 30.958
ATOM	632	0	CYS	Α	67	-10.393	5.716	30.356
ATOM	633	CB		A	67	-8.251	4.003	
ATOM	634	SG	CYS		67	-7.170	2.529	30.217 29.499
ATOM	635	N	GLY		68	-11.877	4.947 4.286	28.791
MOTA	636	H	GLY		68	-12.125	5.984	29.903
ATOM	637	CA	GLY		68	-12.788 -12.581	7.322	29.241
ATOM	638	С	GLY		68	-12.581	8.253	29.376
ATOM	639	0	GLY		68 69	-11.504	7.545	28.471
ATOM	640	N		A	69	-10.817	6.827	28.360
ATOM	641	H		A	69	-11.305	8.800	27.793
ATOM	642 643	CA C		A A	69	-11.838	8.679	26.399
ATOM	644	0	HIS	A	69	-11.516	7.742	25.630
ATOM ATOM	645	CB		A	69	-9.831	9.128	27.724
ATOM	646	CG	HIS	A	69	-9.276	9.286	29.081
MOTA	647	ND1	HIS	A	69	-9.317	10.484	29.778
MOTA	648	HD1		Α	69	-9.688	11.347	29.436
ATOM	649	CD2		A	69	-8.723	8.352	29.912
ATOM	650	CE1		Α	69	-8.783	10.254	30.947
ATOM	651	NE2		A	69	-8.405	8.990	31.091
ATOM	652	N		A	70	-12.768	9.561	25.973
ATOM	653	H		Α	70	-13.084	10.284	26.588
ATOM	654	CA	LYS		70	-13.325	9.492	24.646
ATOM	655	С	LYS		70	-12.346	10.074	23.653
ATOM	656	0	LYS		70	-11.587	11.055	23.864
ATOM	657	CB	LYS	Α	70	-14.645	10.285	24.536
ATOM	658	CG	LYS	Α	70	-15.837	9.703	25.330
ATOM	659	CD	LYS	Α	70	-17.105	10.593	25.286
ATOM	660	CE	LYS	Α	70	-18.293	10.011	26.092
ATOM	661	NZ	LYS	A	70	-18.802	8.702	25.608
ATOM	662	1HZ	LYS		70	-19.563	8.406	26.185
ATOM	663	3HZ	LYS		70	-18.069	8.023	25.650
MOTA	664	2HZ	LYS		70	-19.116	8.795	24.663
MOTA	665	M	ALA		71	-12.323	9.485	22.446
ATOM	666	H	ALA		71	-12.813	8.625	22.305
ATOM	667	CA	ALA		71	-11.616	10.044	21.333
MOTA	668	C	ALA		71	-12.529	9.795	20.171
ATOM	669	0	ALA		71	-13.351	8.850	20.146
ATOM	670	CB	ALA		71	-10.292	9.358	21.143
ATOM	671	N	ILE	Α	72	-12.559	10.685	19.149

FIG. ILL SUBSTITUTE SHEET (RULE 26)

					_		
ATOM	672	Н	ILE A	72	-12.006	11.517	19.200
ATOM	673	CA	ILE A			10.474	17.963
ATOM	674	C	ILE A			10.662	16.771
ATOM	675	0	ILE A	72		11.720	16.550
ATOM	676	CB	ILE A	72	-14.541	11.464	17.882
ATOM	677	CG			-15.306	11.455	19.196
MOTA	678	CG2				11.203	
ATOM	679	CD1			-16.446	12.415	19.176
ATOM	680	N	GLY A	73	-12.252	9.633	15.958
ATOM	681	H	GLY A	73	-12.778	8.789	16.067
ATOM	682	CA	GLY A	73	-11.253	9.755	14.938
ATOM	683	С	GLY A	73	-11.283	8.554	14.034
ATOM	684	0	GLY A	73	-12.211	7.706	14.006
ATOM	685	N	THR A	74	-10.247	8.428	13.182
MOTA	68.6	H	THR A	74	-9.471	9.055	13.250
ATOM	687	CA	THR A	74	-10.201	7.416	12.158
ATOM	688	C	THR A	74	-9.674	6.134	12.760
ATOM	689	0	THR A	74	-8.670	6.034	13.497
ATOM	690		. THR A	74	-9.298	7.895	11.048
MOTA	691	OG1		74	-9.910	9.019	10.441
MOTA	692	HGl		74	-9.335	9.362	9.698
ATOM	693	CG2		74	-9.088	6.823	9.946
MOTA	694	N	VAL A	75	-10.318	5.027	12.327
MOTA	695	H	VAL A	75	-11.066	5.114	11.669
ATOM	696	CA	VAL A	75	-9.968	3.717	12.778
ATOM	697	C	VAL A	75	-9.906	2.843	11.551
MOTA	698	0	VAL A	75	-10.803	2.807	10.681
ATOM	699	CB	VAL A	75	-11.044	3.250	13.737
ATOM	700	CG1	VAL A	75	-11.021	1.721	13.943
ATOM	701	CG2	VAL A	75	-10.915	4.019	15.034
MOTA	702	N	LEU A	76	-8.768	2.139	11.366
ATOM	703	H	LEU A	76	-8.002	2.260	11.998
ATOM	704	CA	LEU A	76	-8.566	1.183	10.276
ATOM	705	С	LEU A	76		-0.211	10.808
MOTA	706	0	LEU A	76		-0.582	11.958
ATOM	707	CB	LEU A	76	-7.103	1.270	9.798
ATOM	708	CG	LEU A	76	-6.608	2.684	9.443
ATOM ATOM	709 710	CD1	LEU A	76	-5.151	2.645	9.087
ATOM	711	CD2	LEU A VAL A	76	-7.396 2.560	3.302	8.296
ATOM	712			77 77		-1.062	10.042
ATOM	713	H CA	VAL A VAL A	77 77		-0.766 -2.428	9.144 10.485
ATOM	714	C	VAL A	77		-3.412	9.482
ATOM	715	0	VAL A	77		-3.412	8.253
ATOM	716	CB	VAL A	77		-2.592	10.506
ATOM	717	CG1	VAL A	77		-4.021	10.682
ATOM	718	CG2	VAL A	77		-1.765	11.634
ATOM	719	N	GLY A	78		-4.402	9.928
ATOM	720	H	GLY A	78		-4.402 -4.530	10.913
ATOM	721	CA	GLY A	78		-5.285	8.987
ATOM	722	C	GLY A	78		-6.380	9.732
ATOM	723	Õ	GLY A	78		-6.524	10.970
ATOM	724	N	PRO A	79		7.271	9.003
ATOM	725	CA	PRO A	79		8.467	9.602
ATOM	726	C	PRO A	79		8.107	10.340
ATOM	727	ō	PRO A	79	_	8.489	10.032
				-	·		

FIG. I IM SUBSTITUTE SHEET (RULE 26)

				2	7/46			
ATOM	728	СВ	PRO	A	79	-5.613	-9.379	8.400
ATOM	729	CG	PRO	Α	79	-5.529	-8.416	7.210
MOTA	730	CD	PRO	Α	79	-6.415	-7.225	7.537
ATOM	731	N	THR	Α	80	-4.759	-7.304	11.408
ATOM	732	H	THR	A	80	-5.664	-6.935	11.619
ATOM	733	CA	THR	Α	80	-3.658	-6.957	12.263
ATOM	734	С	THR	Α	80	-3.490	-8.075	
MOTA	735	0	THR	Α	80	-4.447	-8.642	13.857
ATOM	736	CB	THR	Α	80	-3.868	-5.572	12.927
MOTA	737	OG1	THR		80	-2.770	-5.303	13.787
MOTA	738	HG1	THR	Α	80	-2.889	-4.412	14.225
ATOM	739	CG2	THR		80	-5.210	-5.464	13.678
ATOM	740	N	PRO		81	-2.243	-8.496	13.589
ATOM	741	CA	PRO		81	-1.986	-9.476	14.660
ATOM	742	С		A	81	-2.499	-8.952	16.001
ATOM	743	0		A	81	-2.944	-9.720	16.866
ATOM	744	CB		A	81	-0.444	-9.549	14.732
ATOM	745	CG	PRO		81	0.069	-8.951	13.429
ATOM	746	CD	PRO		81	-1.029	-8.105	12.842
ATOM	747	N	VAL		82	-2.474	-7.621	16.276
ATOM	748	H	VAL		82	-2.180	-6.975	15.571 17.591
MOTA	749	CA.	VAL		82	-2.869	-7.091 -5.761	17.391
ATOM	750	С	VAL		82	-3.605 -3.349	-5.761 -5.004	16.429
ATOM	751	O	VAL		82	-3.349	-6.858	18.443
ATOM	752	CB	VAL VAL		82 82	-0.650	-5.824	17.803
ATOM	753	CG1	VAL		82	-1.907	-6.418	19.890
ATOM ATOM	754 755	CG2 N	ASN		83	-4.548	-5.371	18.260
ATOM	756	H	ASN		83	-4.810	-5.981	19.007
ATOM	757	CA	ASN		83	-5.181	-4.067	18.123
ATOM	758	C	ASN		83	-4.195	-3.019	18.565
ATOM	759	Õ	ASN		83	-3.605	-3.064	19.665
ATOM	760	CB	ASN		83	-6.436	-3.942	18.982
ATOM	761	CG	ASN		83	-7.502	-4.930	18.631
ATOM	762	OD1	ASN		83	-7.899	-5.049	17.488
ATOM	763	ND2	ASN		83	-7.980	-5.662	19.628
ATOM	764	2HD2		Α	83	-8.695	-6.341	19.459
ATOM	765	1HD2	ASN	Α	83	-7.630	-5.541	20.557
ATOM	766	N	ILE	Α	84	-4.007	-1.951	17.770
ATOM	767	H	ILE	Α	84	-4.583	-1.827	16.962
MOTA	768	CA	ILE		84	-2.993	-0.954	18.032
ATOM	769	С	ILE		84	-3.679	0.387	18.114
ATOM	770	0	ΙĻΕ		84	-4.460	0.797	17.240
ATOM	771	CB	ILE		84	-2.021	-0.922	16.833
ATOM	772	CG1		A	84	-1.162	-2.150	16.859
ATOM	773.		ILE		84	-1.219	0.387	16.747
ATOM	774	CD1	ILE		84	-0.375	-2.360	15.579
ATOM	775	N	ILE		85	-3.471	1.155	19.203
ATOM	776	H		A	85 85	-2.972	0.781	19.985
ATOM	777	CA		A R	85 85	-3.951 -2.784	2.518	19.281 18.949
ATOM	778	C		A A	85 85	-2.784 -1.767	3.425 3.515	19.663
ATOM	779	O		A A	85 85	-4.522	2.825	20.676
ATOM	780	CB CG1		A	85	-5.673	1.865	21.050
ATOM	781 782	CG1 CG2		A	85	-5.000	4.274	20.716
ATOM ATOM	783	CD1	ILE		85	-6.828	1.808	20.059
*** 01.1	, , ,	-171				3.023		

FIG. IN SUBSTITUTE SHEET (RULE 26)

2	R	1	L.	6
_	.,	•	•	

ATOM	78	4 N	GLY	A 86	-2.820	4.123	17.792
ATOM	78	5 H	GLY	_		4.087	
ATOM	78					4.936	
ATOM	78			A 86		6.393	
ATOM	788		GLY		-2.760	6.864	
ATOM	789		ARG		-0.881		
ATOM	790		ARG A			7.229	
ATOM	791				-0.204	6.890	
ATOM	. 792		ARG A		-0.810	8.623	17.643
ATOM	793				-2.027	9.445	17.277
ATOM			ARG A		-2.365	10.430	17.963
ATOM	794				0.450	9.275	17.057
	795				1.735	8.496	17.205
ATOM	796				2.762	8.916	16.207
ATOM	797				3.875	7.961	16.117
ATOM	79.8				4.035	7.353	16.895
ATOM	799				4.660	7.893	15.035
ATOM	800				4.463	8.675	13.975
ATOM	801	2HH			3.712	9.335	13.974
ATOM	802	1HH			5.066	8.602	13.181
ATOM	803	NH:		. 87	5.656	7.019	15.023
ATOM	804	1HH:	2 ARG A	. 87	6.254	6.953	14.224
ATOM	805	2HH:	2 ARG A	87	5.810	6.426	15.813
ATOM	806	N	ASN A	88	-2.780	9.120	16.214
ATOM	807	H	ASN A		-2.504	8.361	15.625
ATOM	808	CA	ASN A		-4.015	9.860	15.890
ATOM	809	С	ASN A		-4.963	9.921	17.069
· · ATOM	810	0	ASN A	88	-5.613	10.954	17.345
ATOM ·	811	CB	ASN A	88	-4.712	9.315	14.617
ATOM	812	CG	ASN A	88	-5.475	8.001	14.827
ATOM	813	OD1		88	-4.922	6.996	15.245
ATOM	814	ND2		88	-6.758	7.998	14.506
ATOM	815	2HD2		88	-7.306	7.169	14.622
ATOM	816	1HD2		88	-7.190	8.824	
ATOM	817	N	LEU A	89	-5.130	8.847	14.145
ATOM	818	H	LEU A	89	-4.637	8.002	17.848
ATOM	819	CA	LEU A	89	-6.024		17.640
ATOM	820	c.	LEU A	89	-5.275	8.865	19.013
ATOM	821	Õ	LEU A	89		9.091	20.309
ATOM	822	СВ	LEU A	89	-5.834	9.632	21.283
ATOM	823	CG	LEU A	89	-6.840	7.592	19.140
ATOM	824	CD1		89	-7.759	7.355	17.957
ATOM	825	CD2		89	-8.369	5.980	18.088
ATOM	826	N N	LEU A		-8.817	8.457	17.801
ATOM	827		LEU A	90	-3.983	8.745	20.428
ATOM	828	H CA		90	-3.525	8.274	19.674
ATOM	829		LEU A	90	-3.242	9.057	21.664
ATOM		C	LEU A	90		10.555	21.932
MOTA	830	0	LEU A	90		1.020	23.092
	831	CB	LEU A	90	-1.817	8.453	21.661
ATOM	832	CG	LEU A	90	-1.766		21.587
ATOM	833	CD1	LEU A	90.	-0.343		21.396
ATOM	834	CD2	Leu a	90	-2.339	6.230	22.812
ATOM	835	Ŋ	THR A	91	-3.031 1		20.926
ATOM	836	H	THR A	91		1.063	19.988
ATOM	837	CA	THR A	91	-2.964 1		21.155
ATOM	838	С	THR A	91	-4.309 1		21.635
ATOM	839	0	THR A	91			22.398

13.543 19.848 91 -2.555 THR A MOTA 840 CB 13.214 18.802 -3.459 91 MOTA 841 OG1 THR A 17.958 -3.188 91 13.677 **ATOM** 842 HG1 THR A 13.122 19.395 -1.153 91 CG2 THR A MOTA 843 12.704 21.258 -5.435 N GLN A 92 MOTA 844 20.677 11.892 -5.379 GLN A 92 MOTA 845 Н 13.186-21.682 -6.763 GLN A 92 MOTA 846 CA 12.975 23.153 -6.942 GLN A 92 MOTA 847 C -7.554 13.797 23.871 **ATOM** 848 0 GEN A 92 -7.890 20.964 12.479 **ATOM** 849 CB GLN A 92 -7.937 12.862 19.517 850 CG GLN A 92 **ATOM** -9.251 12.515 18.886 **ATOM** 851 CD GLN A 92 19.546 -10.270 12.424 852 OE1 GLN A 92 **ATOM** 17.588 NE2 GLN A 92 -9.202 12.323 853 MOTA -10.031 12.087 17.080 854 1HE2 GLN A 92 MOTA -8.336 12.411 17.097 **ATOM** 855 2HE2 GLN A 92 -6.472 11.846 23.721 **ATOM** 856 N ILE A 93 -6.014 11.160 23.155 MOTA ILE A 93 857 H -6.608 11.578 25.165 ILE A 93 **ATOM** 858 CA -5.472 12.189 25.948 **ATOM** C ILE A 93 859 12.031 27.171 -5.342 0 ILE A 93 MOTA 860 10.073 25.484 -6.820 CB ILE A 93 MOTA 861 9.221 25.286 CG1 ILE A -5.536 93 **ATOM** 862 9.486 24.735 -8.022 CG2 ILE A 93 MOTA 863 -5.754 7.740 25.693 **ATOM** 864 CD1 ILE A 93 12.993 25.330 GLY A -4.594 865 N 94 MOTA 24.334 13.079 GLY A -4.617 94 MOTA 866 H 13.742 26.063 -3.613 GLY A MOTA 867 CA 94 26.512 12.895 -2.448 **ATOM** 868 C GLY A 94 13.158 27.519 -1.764869 0 GLY A 94 MOTA -2.117 11.849 25.797 CYS A **MOTA** 870 N 95 11.644 24.957 CYS A -2.619 **ATOM** 871 H 95 -1.036 10.994 26.214 **ATOM** 872 CA CYS A 95 11.566 25.925 0.362 **ATOM** 873 С CYS A 95 0.588 12:254 24.907 ATOM 874 0 CYS A 95 9.655 25.550 -1.260 MOTA 875 CB CYS A 95 8.307 26.125 CYS A 95 -0.254MOTA 876 SG 11.297 26.803 THR A 96 1.346 MOTA 877 N 1.135 10.738 27.618 THR A MOTA 878 Н 96 2.728 11.779 26.664 THR A 96 **ATOM** 879 CA 3.729 10.784 27.264 THR A 96 **ATOM** 880 C 10.249 THR A 3.498 28.345 96 **ATOM** 881 0 2.925 13.154 27.346 THR A 96 MOTA 882 CB 2.594 13.109 28.721 OG1 THR A 96 **ATOM** 883 13.966 2.784 29.109 HG1 THR A 96 MOTA 884 14.300 26.698 2.139 CG2 THR A 96 **ATOM** 885 26.599 4.882 10.603 LEU A 97 MOTA 886 N 25.714 5.016 11.071 LEU A 97 ATOM 887 Н 27.166 6.040 9.910 **ATOM** CA LEU A 97 888 6.751 10.824 28.175 LEU A 97 ATOM C 889 12.046 6.705 28.044 LEU A 97 ATOM 890 0 26.049 7.013 9.497 97 LEU A ATOM 891 CB 25.065 6.452 8.449 LEU A 97 **ATOM** 892 CG 8.355 23.828 97 7.360 CD1 LEU A MOTA 893 6.345 7.065 25.724 97 CD2 LEU A **ATOM** 894 7.412 10.221 29.175 98 ASN A MOTA 895 N

FIG. IP
SUBSTITUTE SHEET (RULE 26)

				3	30/46			
ATOM	896	н	ASN	Α	98	7.413	9.212	29.205
ATOM	897	CA	ASN	Α	98	8.065	10.897	30.292
ATOM	898	С	ASN	Α	98	9.220	10.029	30.800
ATOM	899	Ō	ASN	Α	98	8.995	9.079	31.550
ATOM	900	CB	ASN	A	98	7.057	11.177	31.423
ATOM	901	CG	ASN	A	98	6.084	12.305	31.083
ATOM	902	OD1	ASN	A	98	4.983	12.062-	30.594
ATOM	903	ND2	ASN	A	98	6.493	13.549	31.342
ATOM	904	2HD2	ASN	A	98	5.888	14.331	31.136
ATOM	905	1HD2	ASN	Α	98	7.406	13.707	31.742
ATOM	906	N	LEU	A	99	10.451	10.369	30.389
MOTA	907	H	LEU	Α	99	10.547	11.177	29.792
ATOM	908	CA		A	99	11.679	9.620	30.666
ATOM	909	C	LEU	A	99	12.711	10.437	31.454
ATOM	910	Ö		A	99	12.487	11.652	31.651
ATOM	911	CB	LEU	Α	99	12.233	8.989	29.369
ATOM	912	CG	LEU	A	99	12.833	9.873	28.248
ATOM	913	CD1	LEU	A	99	11.876	10.947	27.705
ATOM	914	CD2	LEU	A	99	14.183	10.505	28.623
ATOM	915	OXT		Α	99	13.716	9.819	31.869
TER	710	0				•	·	
ATOM	916	N	PRO	В	1	12.600	14.237	30.106
ATOM	917	CA	PRO	В	1	11.842	15.268	29.363
ATOM	918	C	PRO	В	1	10.430	14.773	29.138
ATOM	919	Ō	PRO	В	1	10.054	13.695	29.618
ATOM	920	CB	PRO	В	1	12.622	15.412	28.035
ATOM	921	CG	PRO	В	1	13.817	14.470	28.131
ATOM	922	CD	PRO	В	1	13.966	14.227	29.603
ATOM	923	1H	PRO	В	1	12.175	13.343	29.964
ATOM	924	2H	PRO	В	1	12.594	14.457	31.081
ATOM	925	N	GLN	В	2	9.513	15.542	28.523
MOTA	926	H	GLN	В	2	9.751	16.474	28.251
MOTA	927	CA	GLN		2 .	8.186	15.058	28.242
ATOM	928	C	GLN	В	2	8.066	15.151	26.749
MOTA	929	0	GLN	В	2	8.523	16.140	26.133
ATOM	930	CB	GLN	В	2	7.155	15.976	28.856
MOTA	931	CG	GLN	В	2	5.739	15.732	28.373
ATOM	932	CD	GLN	В	2	4.744	16.365	29.284
MOTA	933	OE1		В	2	4.628	15.962	30.431
ATOM	934	NE2	GLN	В	2	4.024	17.367	28.784
ATOM	935	1HE2		В	2	3.341	17.830	29.349
ATOM	936	2HE2		В	2	4.160	17.665	27.839 26.036
ATOM	937	N	ILE	В	3	7.499	14.176	26.504
ATOM	938	H	ILE	В	3	7.102	13.386	24.601
ATOM	939	CA	ILE	В	3	7.435	14.216	24.184
ATOM	940	C	ILE	B	3	5.956	14.097 13.290	24.710
ATOM	941	0	ILE	В	3	5.150	13.250	24.029
ATOM	942	CB	ILE	В	3 3	8.299 9.743	13.036	24.534
ATOM	943	CG1	ILE	В	3	8.269	12.985	22.496
ATOM	944	CG3	ILE	B	3	10.621	12.068	24.143
ATOM	945	CD1	ILE	8		5.462	15.108	23.453
ATOM	946	N	THR THR	8	4 4	6.046	15.887	23.226
ATOM	947	H CA		B B	4	4.107	15.115	22.976
ATOM ATOM	948 949	CA		B	4	4.039	14.193	21.765
ATOM	950	0	THR		4	5.066	13.755	21.203
MI OU	7 3 0	_		_	-			

FIG. I IQ SUBSTITUTE SHEET (RULE 26)

				31/46	5		
ATOM	95:	1 CB	THR	B 4	3.616	16.548	22.647
MOTA	952			B 4	4.450	17.157	21.645
ATOM	953			B 4	4.123	18.080	21.442
ATOM	954			B 4	3.644	17.454	23.876
ATOM	959			B 5	2.872	13.781	21.324
ATOM	956			B 5	2.033	14.151	21.723
ATOM	957			B 5	2.837	12.795	
ATOM	958			В 5	2.183	13.415	19.047
ATOM	959			3 5	1.677	12.720	18.142
ATOM	960	CB		3 5	2.093	11.577	20.762
ATOM	961	. CG	LEU I		2.819	10.856	21.892
ATOM	962	CD:	L LEU F	3 5	1.889	9.885	22.602
ATOM	963	CD	LEU F	3 5	4.108	10.159	21.416
ATOM	964	N	TRP F	3 6	2.209	14.742	18.880
ATOM	965		TRP E		2.601	15.323	19.593
ATOM	966		TRP E		1.683	15.364	17.690
ATOM	967		TRP E		2.581	14.978	16.509
ATOM	968		TRP E		2.159	14.851	15.349
ATOM	969		TRP E		1.587	16.879	17.833
ATOM	970		TRP E		0.652	17.339	18.921
ATOM	971				0.955	17.584	20.232
ATOM	972				-0.750	17.612	18.783
ATOM	973			_	-0.167	17.989	20.913
ATOM	974			-	-0.217	18.230	21.882
ATOM	975				-1.224	18.013	20.048
ATOM ATOM	976 977				-1.637	17.550	17.709
ATOM	977 978	CZ2		_	-2.544	18.352	20.266
ATOM	979	CH2		_	-2.947	17.885	17.921
ATOM	980	N N	GLN B	_	-3.394 3.896	18.281	19.185 16.738
ATOM	981	H	GLN B	7	4.267	14.809 14.985	17.650
ATOM	982	CA	GLN B	7	4.794	14.376	15.689
ATOM	983	C.	GLN B	7	5.361	13.043	16.096
ATOM	984	Ō	GLN B	7	5.221	12.586	17.243
ATOM	985	CB	GLN B	7	5.880	15.430	15.505
ATOM	986	CG	GLN B	7	5.353	16.704	14.804
ATOM	987	CD	GLN B	7	6.197	17.912	15.137
ATOM	988	OE1	GLN B	7	7.400	17.802	15.404
ATOM	989	NE2	GLN B	7	5.553	19.083	15.121
ATOM	990	1HE2	GLN B	7	6.040	19.931	15.330
ATOM	991	2HE2	GLN B	7	4.579	19.121	14.900
ATOM	992	N	ARG B	8	5.979	12.274	15.189
ATOM	993	H	ARG B	8	6.073	12.597	14.247
ATOM	994	CA	ARG B	8	6.505	10.985	15.573
ATOM	995	C	ARG B	8	7.577	11.198	16.610
ATOM	996	0	ARG B	8	8.395	12.130	16.515
ATOM	997	CB	ARG B	8	7.092	10.238	14.384
ATOM ATOM	998	CG	ARG B	8	6.132	10.018	13.237
ATOM	999 1000	CD	ARG B	8	6.802	9.4.02	12.046
ATOM	1000	ne He	ARG B	8 8	5.846	9.005	11.023
ATOM	1001	CZ	ARG B	8	4.872	9.080	11.237
ATOM	1002	NH1	ARG B	8	6.217	8.552	9.828
ATOM	1003		ARG B	8	7.496 8.211	8.442 8.703	9.486
ATOM	1005	1HH1	ARG B	8	7.744	8.703	10.134
ATOM	1006	NH2	ARG B	8	5.279	8.202	8.580 8.952
_				_	3.213	J. 202	J. 932

FIG. 1 IR SUBSTITUTE SHEET (RULE 26)

				3	2/46			
MOTA	1007	1HH2	ARG	В	8	5.540	7.860	8.050
ATOM	1008	2HH2	ARG	В	8	4.312	8.281	9.196
ATOM	1009	N	PRO	В	9	7.663	10.381	17.682
ATOM	1010	CA	PRO	В	9	8.666	10.587	18.746
ATOM	1011	C	PRO	В	9	10.065	10.196	18.315
ATOM	1012	Õ	PRO	В	9	10.678	9.215	18.778
ATOM	1012	CB	PRO	В	9	8.148	9.682-	19.878
ATOM	1013	CG	PRO	В	9	7.315	8.607	19.206
MOTA	1014	CD	PRO	В	9	6.708	9.323	18.004
ATOM	1015	N	LEU	В	10	10.685	10.969	17.400
ATOM	1017	Н	LEU	В	10	10.201	11.746	16.998
ATOM	1018	CA	LEU	В	10	12.040	10.706	16.978
	1019	C	LEU	В	10	12.976	11.498	17.850
ATOM	1020	0	LEU	В	10	12.880	12.733	18.018
ATOM	1021	СВ	LEU	В	10	12.250	11.170	15.554
ATOM	1021	CG	LEU	В	10	11.427	10.386	14.551
ATOM	1022	CD1	LEU		10	11.385	11.175	13.276
ATOM	1023	CD2	LEU	В	10	11.956	8.947	14.355
ATOM	1024	N CD2	VAL		11	14.030	10.843	18.384
ATOM		H	VAL	В	11	14.148	9.866	18.206
ATOM	1026	CA	VAL	В	11	15.018	11.517	19.223
ATOM	1027	CA	VAL	В	11	16.400	11.111	18.740
ATOM	1028		VAL	В	11	16.581	10.201	17.911
ATOM	1029	O CB		В	11	14.857	11.100	20.699
MOTA	1030		VAL	В	11	13.514	11.586	21.293
MOTA	1031	CG1	VAL	В	11	15.038	9.573	20.903
MOTA	1032	CG2			12	17.485	11.739	19.232
ATOM	1033	N	THR	В	12	17.370	12.507	19.862
ATOM	1034	H	THR THR		12	18.843	11.325	18.868
ATOM	1035	CA		В	12	19.377	10.284	19.837
ATOM	1036	C O		В	12	19.237	10.352	21.082
ATOM	1037			В	12	19.830	12.520	18.820
MOTA	1038	CB	THR	В	12	19.389	13.483	17.876
ATOM	1039	OG1 HG1	THR	В	12	20.028	14.252	17.848
ATOM	1040	CG2	THR	В	12	21.234	12.075	18.399
MOTA	1041		ILE	В	13	20.044	9.234	19.338
ATOM	1042	N H	ILE	В	13	20.135	9.130	18.348
MOTA	1043	CA	ILE	В	13	20.641	8.239	20.176
ATOM	1044 1045		ILE		13	22.119	8.226	19.855
ATOM	1045	C O	ILE		13	22.579	8.817	18.865
ATOM		CB	ILE		13	19.993	6.870	19.879
ATOM	1047	CG1	ILE		13	20.192	6.464	18.415
ATOM	1048	CG2	ILE	В	13	18.482	6.893	20.206
ATOM	1049	CD1	ILE	В	13	19.829	5.035	18.106
ATOM	1050	N	LYS	В	14	22.973	7.618	20.661
ATOM	1051 1052	H	LYS	В	14	22.652	7.243	21.531
ATOM		CA	LYS	В	14	24.364	7.480	20.317
ATOM	1053	CA	LYS	В	14	24.680	6.029	20.477
ATOM	1054 1055	0	LYS	В	14	24.353	5.353	21.484
MOTA			LYS	В	14	25.266	8.263	21.242
ATOM ATOM	1056	CB	LYS	8	14	24.947	9.729	21.236
MOTA	1057	CD	LYS	8	14	25.664	10.498	22.339
ATOM	1058	CE	LYS	В	14	26.758	11.441	21.807
ATOM	1059		LYS	В	14	28.026	10.781	21.440
ATOM	1060	NZ		В	14	28.674	11.466	21.107
MOTA	1061	1HZ				27.855	10.107	20.722
MOTA	1062	3HZ	LYS	0	14	21.033	10.10/	20.722

FIG. 1 IS

				3	33/46			
ATOM	1063	2HZ	LYS	В	14	28.408	10.323	22.243
ATOM	1064	N	ILE	В	15	25.214	5.390	19.425
ATOM	1065	Н	ILE	В	15	25.434	5.901	18.594
ATOM	1066	CA	ILE	В	15	25.489	3.989	19.434
ATOM	1067	C	ILE	В	15	26.832	3.981	18.750
ATOM	1068	0	ILE	В	15	27.104	4.869	17.933
ATOM	1069	CB	ILE	В	15	24.435	3.220	18.606
ATOM	1070	CG1	ILE	В	15	24.893	1.824	18.347
ATOM	1071	CG2	ILE	В	15	24.048	3.977	17.309
ATOM	1072	CD1	ILE	В	15	23.830	0.996	17.645
ATOM	1073	N	GLY	В	16	27.812	3.212	19.202
ATOM	1074	H	GLY	В	16	27.623	2.535	19.913
ATOM	1075	CA		В	16	29.175	3.336	18.677
ATOM	1076	С		В	16	29.771	4.754	18.619
MOTA	1077	0	GLY	В	16	30.737	4.970	17.902
ATOM	1078	N	GLY	В	17	29.273	5.791	19.335
ATOM	1079	H		В	17	28.453	5.660	19.892
ATOM	1080	CA	GLY	В	17	29.924	7.105	19.302
MOTA	1081	C	GLY	В	17	29.468	8.043	18.176
ATOM	1082	0		В	17	29.984	9.155	17.933
ATOM	1083	N		В	18	28.433	7.621	17.411
MOTA	1084	H		В	18	28.046	6.711	17.560
MOTA	1085	CA	_	В	18	27.834	8.449	16.348 16.736
ATOM	1086	C		В	18	26.407	8.755	17.353
ATOM	1087	0		B	18	25.678	7.953 7.645	15.045
ATOM	1088	CB		В	18	27.810 27.247	6.204	15.146
MOTA	1089	CG		В	18	27.572	5.333	13.924
MOTA	1090	CD		В	18	26.771	4.501	13.464
ATOM	1091	OE1 NE2	GLN GLN	B B	18 18	28.766	5.531	13.393
ATOM	1092 1093	1HE2		В	18	29.057	5.005	12.594
ATOM ATOM	1093	2HE2		В	18	29.388	6.209	13.786
ATOM	1095	N		В	19	25.873	9.933	16.337
ATOM	1096	Н		В	19	26.446	10.602	15.863
ATOM	1097	CA		В	19	24.467	10.267	16.578
ATOM	1098	C	LEU	В	19	23.633	9.622	15.490
ATOM	1099	ō		В	19	23.912	9.707	14.284
ATOM	1100	CB	LEU		19	24.207	11.777	16.457
ATOM	1101	CG	LEU		19	24.857	12.756	17.454
ATOM	1102	CD1	LEU		19	24.739	12.335	18.880
ATOM	1103	CD2	LEU	В	19	26.299	13.072	17.130
ATOM	1104	N	LYS	В	20	22.450	9.085	15.850
ATOM	1105	H	LYS	В	20	22.242	8.948	16.819
ATOM	1106	CA	LYS	В	20	21.472	8.702	14.867
MOTA	1107	C		В	20	20.121	9.105	15.417
ATOM	1108	0		В	20	19.957	9.572	16.569
MOTA	1109	CB		В	20	21.496	7.200	14.560
ATOM	1110	CG		В	20	22.904	6.653	14.507
MOTA	1111	CD		B	20	23.052	5.366	13.677 12.145
atom	1112	CE	LYS	B	20	23.069	5.603	12.195
ATOM	1113	NZ	LYS	B	20	23.893	6.758	10.703
ATOM	1114	1HZ		B	20	23.847	6.836 6.617	11.978
ATOM	1115	3HZ		B	20 20	24.843 23.544	7.597	12.116
ATOM	1116	2HZ		B	20 21	19.068	9.022	14.591
ATOM	1117	N H	GLU GLU		21	19.200	8.712	13.650
ATOM	1118	п	GHO			19.200		

FIG. 1 IT SUBSTITUTE SHEET (RULE 26)

				3	4/46			
ATOM	1119	CA	GLU		21	17.735	9.366	15.008
ATOM	1120	C	GLU	В	21	16.937	8.095	15.119
ATOM	1121	0	GLU	В	21	17.117	7.103	14.376
ATOM	1122	CB	GLU	В	21	17.143	10.314	13.983
ATOM	1123	CG	GLU	В	21	15.714	10.706	14.162
ATOM	1124	CD	GLU	В	21	15.304	11.607	13.036
ATOM	1125	OEI		В	21	14.971	11.051	11.957
MOTA	1126	OE2			21	15.338	12.854	13.174
MOTA	1127	N	ALA		22	16.025	7.999	16.072
ATOM	1128	H	ALA		22	15.825	8.792	16.648
ATOM	1129	CA		В	22	15.300	6.783	16.315
MOTA	1130	C		В	22	13.981	7.132	16.952 17.632
ATOM	1131	0		В	22	13.756	8.153 5.865	17.032
ATOM	1132	CB		В	22	16.095	6.230	16.743
MOTA	1133	N	LEU		23 23	12.994 13.195	5.379	16.257
ATOM	1134	H	LEU		23	11.639	6.408	17.180
ATOM	1135	CA C	LEU	В	23	11.476	5.740	18.534
ATOM ATOM	1136 1137	0	LEU		23	11.814	4.564	18.746
ATOM	1138	СВ	LEU	B	23	10.775	5.665	16.192
ATOM	1139	CG	LEU	В	23	9.267	5.810	16.237
ATOM	1140	CD1		В	23	8.807	7.142	15.664
ATOM	1141	CD2		В	23	8.648	4.625	15.482
ATOM	1142	N	LEU	В	24	10.948	6.455	19.553
ATOM	1143	Н	LEU	В	24	10.775	7.433	19.435
ATOM	1144	CA		В	24	10.613	5.838	20.849
ATOM	1145	C	LEU	В	24	9.271	5.160	20.687
ATOM	1146	0	LEU		24	8.208	5.764	20.418
ATOM	1147	CB		В	24	10.564	6.878	21.971
ATOM	1148	CG	LEU	В	24	11.828	7.750	22.075
ATOM	1149	CD1	LEU	В	24	11.580	8.859	23.077
MOTA	1150	CD2		В	24	13.099	6.955	22.388
ATOM	1151	N		В	25	9.246	3.822	20.809
ATOM	1152	H	ASP		25	10.025	3.347	21.218
ATOM	1153	CA		В	25	8.122	3.030	20.366
ATOM	1154	C		В	25	7.637	2.136	21.484
MOTA	1155	0		В	25	8.189	1.048	21.759
MOTA	1156	CB	ASP		25 25	8.613	2.196	19.189 18.511
MOTA	1157	CG	ASP	B	25 25	7.528 6.422	1.421 1.339	19.058
ATOM	1158	OD1 OD2			25	7.800	0.897	17.426
ATOM ATOM	1159 1160	N		В	26	6.547	2.465	22.157
MOTA	1161	H	THR		26	6.067	3.314	21.938
ATOM	1162	CA	THR		26	6.025	1.621	23.212
ATOM	1163	C	THR		26	5.347	0.369	22.694
ATOM	1164	ō	THR		26	4.976	-0.550	23.451
ATOM	1165	CB	THR		26	5.027	2.389	24.046
ATOM	1166	OG1	THR		26	3.927	2.853	23.239
ATOM	1167	HG1	THR		26	3.277	3.359	23.806
ATOM	1168	CG2		В	26	5.703	3.603	24.650
ATOM	1169	N		B	27	5.090	.0.245	21.382
ATOM	1170	H		8	27	5.341	0.983	20.756
ATOM	1171	CA		В	27	4.457	-0.938	20.867
ATOM	1172	С		В	27	5.475	-1.992	20.458
ATOM	1173	0	GLY		27	5.121	-3.108	20.055
MOTA	1174	N	ALA	В	28	6.792	-1.717	20.495

FIG. I IU

			•	35	146			
MOTA	1175	H	ALA	В	28	7.104	-0.832	20.841
MOTA	1176	CA	ALA		28	7.800	-2.690	20.037
ATOM	1177	С	ALA		28	8.371	-3.444	21.259 22.213
MOTA	1178	0	ALA		28	8.840	-2.807	19.358
MOTA	1179	CB	ALA		28	8.924	-1.936	21.289
MOTA	1180	N		В	29	8.459	-4.787	20.535
ATOM	1181	H	ASP		29	8.082	-5.325 -5.441	22.452
MOTA	1182	CA		В	29	9.121	-5.219	22.404
MOTA	1183	C	_	В	29	10.608 11.345	-5.264	23.412
ATOM	1184	0_		В	29	8.965	-6.975	22.447
MOTA	1185	CB		В	29 29	7.551	-7.477	22.774
MOTA	1186	CG		В	29	6.683	-6.693	23.169
MOTA	1187	OD1		B B	29	7.350	-8.686	22.616
ATOM	1188	OD2	ASP ASP		30	11.164	-5.157	21.171
ATOM	1189	N		B	30	10.577	-5.063	20.367
ATOM	1190	H CA		В	30	12.609	-5.217	20.880
ATOM	1191	CA		В	30	13.048	-3.886	20.335
MOTA	1192	Ō		В	30	12.269	-3.055	19.817
ATOM	1193 1194	CB		В	30	12.833	-6.226	19.735
ATOM	1194	CG		В	30	12.477	-7.675	20.099
MOTA MOTA	1196	OD1		В	30	13.197	-8.272	20.908
ATOM	1197	OD2	ASP		30	11.494	-8.237	19.569
ATOM	1198	N		В	31	14.387	-3.692	20.227
MOTA	1199	H	THR		31	15.018	-4.380	20.586
ATOM	1200	CA		В	31	14.981	-2.530	19.614
ATOM	1201	C	THR		31	15.578	-2.979	18.260
ATOM	1202	Ō	THR		31	16.246	-4.020	18.123
MOTA	1203	CB		В	31	16.036	-2.004	20.557
ATOM	1204	OG1	THR	В	31	15.378	-1.376	21.645
ATOM	1205	HG1	THR	В	31	16.052	-1.016	22.290
ATOM	1206	CG2	THR		31	16.944	-0.960	19.904
ATOM	1207	N	VAL		32	15.237	-2.283	17.150 17.237
ATOM	1208	H	VAL		32	14.703	-1.442	
MOTA	1209	CA	VAL		32	15.626	-2.722	15.806 15.132
MOTA	1210	С	VAL		32	16.303	-1.566	14.995
ATOM	1211	0	VAL		32	15.779	-0.428	14.964
ATOM	1212	CB	VAL		32	14.407	-3.126 -3.703	13.596
ATOM	1213	CG1			32	14.820 13.556	-4.102	15.703
ATOM	1214	CG2			32		-1.756	14.720
ATOM	1215	N	LEU		33 33	17.984	-2.658	14.814
ATOM	1216	H	LEU		33	18.347	-0.697	14.138
ATOM	1217	CA C	LEU LEU		33	18.610	-1.009	12.685
ATOM	1218	.0	LEU		33	18.685	-2.162	12.205
ATOM	1219 1220	CB	LEU		33	19.679	-0.628	14.856
ATOM ATOM	1221	CG	LEU		33	19.698	0.363	16.031
ATOM	1222	CD1			33	18.425	0.321	16.891
MOTA	1223	CD2			33	20.929	0.179	16.889
ATOM	1224	N	GLU		34	18.786	0.078	11.899
MOTA	1225	H	GLU		34	18.619	0.991	12.271
ATOM	1226	ĈA	GLU		34	19.218	0.041	10.488
MOTA	1227	Č.	GLU		34	20.478	-0.774	10.399
ATOM	1228	ō	GLU		34	21.374	-0.835	11.272
ATOM	1229	CB	GLU		34	19.536	1.460	9.996
ATOM	1230	CG	GLU		34	20.722	2.088	10.761

FIG. 1 IV SUBSTITUTE SHEET (RULE 26)

WO 01/35316 PCT/US00/30863

				36	/46			
ATOM	1231	CD	GLU	В	34	21.085	3.512	10.314
ATOM	1232	OE1	GLU	В	34	20.285	4.466	10.500
ATOM	1233	OE2	GLU		34	22.211	3.703	9.775
ATOM	1234	N	GLU	В	35	20.673	-1.367	9.205
MOTA	1235	Н	GLU		35	20.011	-1.227	8.468
ATOM	1236	CA		В	35	21.802	-2.205	8.930 9.321
ATOM	1237	C	GLU		35	23.096 23.391	-1.520- -0.379	8.916
ATOM	1238	0	GLU GLU	В	35 35	23.331	-2.479	7.439
ATOM	1239 1240	CB CG	GLU		35	22.795	-3.380	6.883
ATOM ATOM	1241	CD	GLU		35	22.987	-4.587	7.744
ATOM	1242	OE1	GLU		35	21.980	-5.258	8.118
ATOM	1243	OE2	GLU	В	35	24.149	-4.860	8.048
ATOM	1244	N	MET	В	36	23.926	-2.106	10.157
ATOM	1245	H	MET	В	36	23.654	-2.953	10.613 10.441
MOTA	1246	CA	MET	В	36	25.232 26.146	-1.559 -2.687	10.441
MOTA	1247	C	MET	В	36 36	25.731	-3.783	11.257
ATOM	1248	O CB	MET MET	B B	36	25.751	-0.424	11.497
ATOM ATOM	1249 1250	CG	MET	В	36	24.626	-0.724	12.881
ATOM	1251	SD	MET	В	36	24.722	0.719	13.988
ATOM	1252	CE	MET	В	36	23.132	1.586	13.692
ATOM	1253	N	SER	В	37	27.441	-2.551	10.593
ATOM	1254	H	SER		37	27.783	-1.726	10.144 11.011
MOTA	1255	CA	SER		37	28.321	-3.608 -3.352	12.442
MOTA	1256	C	SER		37 37	28.721 29.402	-2.369	12.788
MOTA	1257	O CB	SER SER		3 <i>1</i> 37	29.567	-3.622	10.109
ATOM ATOM	1258 1259	OG	SER		37	29.231	-3.908	8.750
ATOM	1260	HG	SER		37	30.057	-3.911	8.187
ATOM	1261	N	LEU		38	28.469	-4.295	13.366
ATOM	1262	H	LEU		38	27.948	-5.123	13.117
ATOM	1263	CA	LEU		38	29.073	-4.232	14.714 14.895
ATOM	1264	C	LEU		38	30.132 30.070	-5.342 -6.357	14.197
ATOM	1265	0	LEU	В	38 38	27.986	-4.237	15.802
ATOM ATOM	1266 1267	CB CG		В	38	27.005	-3.039	15.750
ATOM	1268	CD1			38	25.885	-3.214	16.788
ATOM	1269	CD2	LEU		38	27.707	-1.696	16.017
ATOM	1270	N	PRO		39	31.119	-5.160	15.804
MOTA	1271	CA	PRO		39	32.199	-6.116	16.052 17.028
MOTA	1272	C	PRO		39	31.767	-7.223 -6.942	18.185
MOTA	1273	0	PRO	В	39 39	31.448 33.347	-5.276	16.625
ATOM	1274 1275	CB CG	PRO PRO	B	39	32.634	-4.148	
ATOM ATOM	1275	CD	PRO		39	31.385	-3.916	16.523
ATOM	1277	N	GLY		40	31.770	-8.481	16.559
ATOM	1278	H	GLY		40	32.036		15.598
ATOM	1279	CA	GLY	В	40	31.420		17.353
ATOM	1280	С	GLY	B	40	30.679		16.539 15.308
atom	1281	0	GLY		40	30.647	-10.671 -11.699	17.255
ATOM	1282	N	LYS	8	41 41	30.098 30.164		18.261
MOTA MOTA	1283 1284	H CA	LYS LYS		41	29.399		16.702
ATOM	1285	CA	LYS		41	27.971		17.245
ATOM	1286	Ö	LYS		41	27.743		18.436

FIG. I IW

17.048 30.154 -14.152 41 LYS B CB 1287 MOTA 31.537 -14.221 16.384 41 CG LYS B 1288 MOTA 32.192 -15.580 16.651 LYS B 41 CD 1289 MOTA 15.983 33.566 -15.642 LYS B 41 CE 1290 MOTA 34.198 -16.956 16.183 LYS B 41 NZ 1291 MOTA 15.732 35.102 -16.968 41 LYS B 1292 1HZ MOTA 15.782 33.612 -17.674 LYS B 41 1293 3HZ MOTA 17.172 34.312 -17.128 41 LYS B 1294 2HZ MOTA 16.351 27.018 -13.228 42 TRP B N 1295 ATOM 15.411 27.307 -13.458 42 TRP B Н 1296 ATOM 16.521 25.597 -12.929 TRP B 42 CA 1297 MOTA 16.405 24.723 -14.179 42 TRP B C MOTA 1298 16.131 25.210 -15.277 42 TRP B ATOM 1299 0 25.192 -11.856 15.491 42 TRP B 1300 CB MOTA 15.390 26.127 -10.687 42 TRP B CG 1301 MOTA 14.244 26.651 -10.197 42 CD1 TRP B 1302 MOTA 16.467 -9.913 26.739 CD2 TRP B 42 1303 MOTA 14.533 -9.191 27.548 NE1 TRP B 42 1304 ATOM 13.818 -8.702 28.067 HE1 TRP B 42 1305 MOTA 15.893 -8.995 27.664 CE2 TRP B 42 1306 MOTA -9.923 17.875 26.640 CE3 TRP B 42 1307 MOTA 16.680 -8.136 28.443 CZ2 TRP B 42 1308 MOTA 18.673 -9.075 27.426 CZ3 TRP B 42 1309 MOTA 18.077 -8.171 28.318 CH2 TRP B 42 1310 MOTA 16.617 23.416 -13.980 43 LYS B N 1311 MOTA 16.840 -13.044 23.105 LYS B 43 Н 1312 MOTA 16.526 22.378 -14.995 LYS B 43 CA 1313 MOTA 21.368 -14.507 15.478 LYS B 43 С 1314 MOTA 15.706 20.743 -13.472 LYS B 43 0 1315 MOTA 21.694 -15.196 17.893 43 LYS B CB 1316 MOTA 19.034 22.641 -15.623 43 LYS B 1317 CG MOTA 22.409 -14.814 20.323 43 LYS B 1318 CD MOTA 20.182 22.767 -13.327 43 LYS B 1319 CE MOTA 20.015 24.214 -13.113 43 LYS B MOTA 1320 NZ 19.924 24.400 -12.125 43 LYS B 1321 1HZ ATOM 24.532 -13.593 19.185 43 LYS B 1322 3HZ MOTA 24.702 -13.476 20.821 LYS B 43 1323 2HZ MOTA 21.175 -15.204 14.341 44 PRO B MOTA 1324 N 13.382 20.139 -14.835 PRO B 44 1325 CA MOTA 14.044 18.765 -14.997 PRO B 44 C 1326 ATOM 18.573 -15.902 14.860 44 PRO B ATOM 1327 0 12.180 20.341 -15.761 44 PRO B CB MOTA 1328 20.999 -16.999 12.787 PRO B 44 1329 MOTA CG 13.933 21.837 -16.434 44 PRO B 1330 CD MOTA 17.825 -14.101 13.712 45 LYS B N MOTA 1331 17.994 -13.483 12.944 LYS B 45 Н MOTA 1332 14.339 16.523 -14.088 45 LYS B 1333 CA MOTA 15.519 -13.590 13.329 LYS B 45 C 1334 ATOM 12.379 15.829 -12.838 45 LYS B 0 MOTA 1335 15.560 16.558 -13.149 LYS B 45 CB MOTA 1336 15.469 -13.442 16.579 45 LYS B CG MOTA 1337 17.501 15.256 -12.254 45 LYS B CD 1338 MOTA 14.131 -12.461 18.469 45 LYS B CE MOTA 1339 19.474 14.549 -13.442 LYS B 45 NZ MOTA 1340 20.126 13.805 -13.588 LYS B 45 1341 1HZ MOTA 19.958 15.355 -13.101 LYS B 45 1342 3HZ MOTA

FIG. IX
SUBSTITUTE SHEET (RULE 26)

			* ***	4-	14.772 -14.306 19.023
MOTA		2HZ	LYS B	45	14.240 -14.005 13.416
MOTA	1344	N	MET B	46	13.991 -14.705 14.085
MOTA	1345	H	MET B	46	13.203 -13.472 12.570
MOTA	1346	CA	MET B	46	12.291 -12.623 13.425
ATOM	1347	C	MET B	46	
ATOM	1348	0	MET B	46	11
ATOM	1349	CB	MET B	46	
ATOM	1350	CG	MET B	46	10.100 ==
ATOM	1351	SD	MET B	46	
ATOM	1352	CE	MET B	46	13.300
ATOM	1353	N	ILE B	47	
ATOM	1354	H	ILE B	47	
ATOM	1355	CA	ILE B	47	
ATOM	1356	С	ILE B	47	
ATOM	1357	0	ILE B	47	
ATOM	13 5 8	CB	ILE B	47	
ATOM	1359	CG1	ILE B	47	
ATOM	1360	CG2	ILE B	47.	
ATOM	1361	CD1	ILE B	47	
ATOM	1362	N	GLY B	48	
ATOM	1363	Н	GLY B	48	8.484 -10.249 14.549
ATOM	1364	CA	GLY B	48	7.365 -9.872 12.800
ATOM	1365	С	GLY B	48	6.826 -8.512 13.141
ATOM	1366	0	GLY B	48	7.136 -7.832 14.149
ATOM	1367	N	GLY B	49	5.940 -8.027 12.306
ATOM	1368	H	GLY B	49	5.668 -8.562 11.506
ATOM	1369	CA	GLY B	49	5.336 -6.745 12.493
ATOM	1370	C	GLY B	49	4.082 -6.786 11.674
ATOM	1371	ō	GLY B	49	3.561 -7.847 11.273
MOTA	1372	N	ILE B	50	3.531 -5.634 11.315
MOTA	1373	H	ILE B	50	4.015 -4.777 11.492
ATOM	1374	CA	ILE B	50	2.247 -5.573 10.673
ATOM	1375	C	ILE B	50	2.118 -6.456 9.420
ATOM	1376	Ö	ILE B	50	1.175 -7.253 9.215
ATOM	1377	CB	ILE B	50	1.982 -4.071 10.391
ATOM	1378	CG1	ILE B	50	1.005 -3.539 11.396
MOTA	1379	CG2		50	1.610 -3.739 8.922
MOTA	1380	CD1		50	-0.391 -4.077 11.252
ATOM	1381	N	GLY B	51	3.113 -6.410 8.519
ATOM	1382	H	GLY B	51	3.957 -5.920 8.737
ATOM	1383	CA	GLY B	51	2.926 -7.075 7.259
MOTA	1384	C	GLY B	51	3.671 -8.391 7.077
ATOM	1385	Ō	GLY B	51	3.716 -8.945 5.973
ATOM	1386	N	GLY B	52	4.296 -8.982 8.116
ATOM	1387	H	GLY B	52	4.227 -8.580 9.029
ATOM	1388	CA	GLY B	52	5.053 -10.190 7.874
ATOM	1389	C	GLY B	52	6.334 -10.178 8.678
ATOM	1390	ō	GLY B	52	6.519 -9.421 9.657
ATOM	1391	Ň	PHE B	53	7.325 -11.015 8.343
ATOM	1392	H	PHE B	53 ⁻	7.227 -11.603 7.540
ATOM	1393	ĊA	PHE B	53	8.542 -11.096 9.110
ATOM	1394	c c	PHE B	53	9.72710.584 8.315
ATOM	1395	Õ	PHE B	53	9.780 -10.618 7.075
ATOM	1396	СВ	PHE B	53	8.804 -12.555 9.542
ATOM	1397	CG	PHE B	53	7.850 -13.023 10.592
ATOM	1398	CD:		53	6.513 -13.277 10.279

FIG. I IY

SUBSTITUTE SHEET (RULE 26)

WO 01/35316 PCT/US00/30863

				3 (9/46	•	
		ana	מעם	·	57 - 0 53	8.279 -13.1	92 11.918
ATOM	1399 1400	CD2 CE1	-	B B	53	5.620 -13.6	97 11.253
MOTA MOTA	1400	CE2		В	53	7.382 -13.6	
ATOM	1402	CZ		В	53	6.052 -13.8	
ATOM	1403	N		В	54	10.758 -10.1	
ATOM	1404	H		В	54	10.665 -9.9	
ATOM	1405	CA	ILE		54	12.029 -9.9 13.089 -10.6	
MOTA	1406	С		В	54	13.089 -10.6 12.952 -11.0	
MOTA	1407	0	ILE		54 54	12.390 -8.4	
MOTA	1408	CB	ILE	B B	5 4 54	12.386 -7.7	
MOTA	1409	CG1 CG2	ILE		54	11.460 -7.7	7.218
MOTA MOTA	1410 1411	CD1	ILE		54	13.113 -6.4	
ATOM	1412	N	LYS		55	14.272 -10.8	
ATOM	1413	H	LYS		55	14.383 -10.5	
ATOM	1414	CA	LYS	В	55	15.403 -11.4	
ATOM	1415	С	LYS		55	16.274 -10.3	
MOTA	1416	0	LYS		55	16.620 -9.3 16.222 -12.2	•
MOTA	1417	CB	LYS		55	15.638 -13.5	_
ATOM	1418	CG		В	55 55	16.299 -14.3	
MOTA	1419	CD CE	LYS LYS		55	15.311 -14.5	
MOTA	1420 1421	NZ		В	55	15.757 -15.5	77 4.897
ATOM ATOM	1421	1HZ		В	55	15.095 -15.6	
MOTA	1423	3HZ	LYS		55	15.830 -16.4	
ATOM	1424	2HZ	LYS	В	55	16.650 -15.3	
ATOM	1425	N	VAL		56	16.880 -10.5	
ATOM	1426	H	VAL		56	16.741 -11.4 17.732 -9.5	
ATOM	1427	CA	VAL		56	17.732 -9.5 18.884 -10.3	
ATOM	1428	C	VAL VAL		56 56	18.884 -11.5	
ATOM	1429 1430	O CB	VAL		56	16.912 -8.8	319 12.609
MOTA MOTA	1431	CG1		В	56	15.865 -7.9	
MOTA	1432	CG2	VAL		56	16.215 -9.7	
ATOM	1433	N	ARG	В	57	19.958 -9.5	
ATOM	1434	H	ARG		57	20.030 -8.6	
ATOM	1435	CA		В	57	21.050 -10.3 20.963 -9.6	
ATOM	1436	C	ARG		57 53	20.963 -9.0	
MOTA	1437	O.	ARG		57 57	22.426 -9.8	
ATOM	1438	CB CG	ARG ARG	В	57	22.664 -10.4	137 11.439
MOTA MOTA	1439 1440	CD		В	57	24.012 -10.0	065 10.899
ATOM	1441	NE		В	57	24.280 -10.0	
ATOM	1442	HE		В	57	23.592 -11.	
ATOM	1443	CZ		В	57	25.392 -10.4	
ATOM	1444	NHl	-	В	57	26.337 -9.0 26.223 -9.0	
MOTA	1445			В	57 53	26.223 -9. 27.163 -9.	
ATOM	1446				57 57	25.561 -11.	
MOTA	1447				57	26.392 -10.	950 7.225
atom atom	1448 1449			B	57	24.857 -11.	729 7.422
atom	1450				58	20.997 -10.	
ATOM	1451		GLN		58	21.176 -11.	
MOTA	1452	CA	GLN		58	20.780 -10.	
ATOM	1453	С	GLN		58	22.108 -9. 22.918 -10.	
MOTA	1454	0	GLN	В	58	22.918 -10.	J_J

FIG. 1 IZ

ATOM 1455 CB GLN B 58	40/46										
ATOM 1456 CG GLN B 58	λ TΩM	1455	CB	GI.N	В	58		20.	051	-11.190	
ATOM 1457 CD GLN B 58											
ATOM 1458 OE1 GLN B 58										-12.003	20.112
ATOM 1459 NE2 GLN B 58								19.	712	-12.472	
ATOM 1460 1HE2 GLN B 58										-12.476	19.623
ATOM 1461 2HE2 GLN B 58										-13.249	20.063
ATOM 1462 N TYR B 59										-12.066	
ATOM 1463 H TYR B 59 21.788 -7.921 18.311 ATOM 1464 CA TYR B 59 23.631 -8.486 19.161 ATOM 1465 C TYR B 59 23.244 -8.290 20.607 ATOM 1466 O TYR B 59 22.178 -7.728 20.927 ATOM 1466 CG TYR B 59 24.387 -7.241 18.653 ATOM 1466 CG TYR B 59 24.387 -7.241 18.653 ATOM 1468 CG TYR B 59 24.387 -7.241 18.653 ATOM 1469 CD1 TYR B 59 24.387 -7.242 16.494 ATOM 1470 CD2 TYR B 59 25.385 -6.753 16.374 ATOM 1471 CE1 TYR B 59 25.385 -6.753 16.374 ATOM 1472 CE2 TYR B 59 25.391 -7.093 15.112 ATOM 1472 CE2 TYR B 59 25.291 -6.603 14.995 ATOM 1473 CT TYR B 59 24.068 -6.774 14.365 ATOM 1474 OH TYR B 59 24.068 -6.774 14.365 ATOM 1475 HH TYR B 59 24.018 -6.6394 12.658 ATOM 1476 N ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 23.644 -8.624 22.992 ATOM 1481 CB ASP B 60 23.644 -8.624 22.992 ATOM 1482 CG ASP B 60 23.777 ATOM 1482 CG ASP B 60 23.777 ATOM 1483 OD1 ASP B 60 23.777 ATOM 1484 CD ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 23.252 -7.234 25.146 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1488 C GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 26.560 -6.763 26.028 ATOM 1490 CB GLN B 61 26.560 -6.765 29.992 ATOM 1491 CG GLN B 61 26.560 -7.675 29.992 ATOM 1493 OEI GLN B 61 26.560 -7.675 29.992 ATOM 1494 NE2 GLN B 61 26.560 -7.675 29.992 ATOM 1499 CD GLN B 61 26.5714 -7.766 28.185 ATOM 1499 CD GLN B 61 26.5714 -7.766 28.185 ATOM 1499 CD GLN B 61 26.5714 -7.766 28.185 ATOM 1499 CD GLN B 61 26.570 -7.675 29.992 ATOM 1499 CD GLN B 61 26.5714 -7.844 29.014 ATOM 1499 CD GLN B 61 26.5714 -7.844 29.014 ATOM 1499 CD GLN B 61 26.5714 -7.844 29.014 ATOM 1499 CD GLN B 61 26.763 -8.038 26.753 ATOM 1499 CD GLN B 61 26.5714 -7.844 29.014 ATOM 1499 CD GLN B 61 26.5714 -7.844 29.014 ATOM 1499 CD GLN B 61 26.660 -7.675 29.992 ATOM 1499 CD GLN B 61 26.660 -7.675 29.992 ATOM 1499 CD GLN B 61 26.764 -8.073 28.669 ATOM 1499 CD GLN B 61 26.2430 -7.675 29.992 ATOM 1499 CD GLN B 61 26.403 -7.022 21.912 ATOM 1490 CD GL								22.	416		
ATOM 1465 C TYR B 59 23.631 -8.486 19.161 ATOM 1465 C TYR B 59 23.244 -8.290 20.607 ATOM 1466 O TYR B 59 22.178 -7.28 20.927 ATOM 1466 C TYR B 59 22.178 -7.241 18.653 ATOM 1468 CG TYR B 59 24.387 -7.241 18.653 ATOM 1469 CD1 TYR B 59 23.045 -7.242 16.494 ATOM 1470 CD2 TYR B 59 23.045 -7.242 16.494 ATOM 1471 CE1 TYR B 59 22.939 -7.093 15.112 ATOM 1472 CE2 TYR B 59 22.939 -7.093 15.112 ATOM 1473 CZ TYR B 59 24.066 -6.774 14.365 ATOM 1474 OH TYR B 59 24.066 -6.774 14.365 ATOM 1475 HH TYR B 59 24.068 -6.620 13.010 ATOM 1475 HH TYR B 59 24.926 -6.394 12.658 ATOM 1476 N ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 23.644 -8.624 22.992 ATOM 1478 CA ASP B 60 23.644 -8.624 22.992 ATOM 1480 O ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 22.803 -10.960 23.332 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 22.803 -10.960 23.332 ATOM 1484 OD2 ASP B 60 22.803 -10.960 23.332 ATOM 1485 N GLN B 61 23.208 -12.126 23.733 ATOM 1486 H GLN B 61 23.208 -12.126 23.733 ATOM 1488 C GLN B 61 25.511 -6.086 25.519 ATOM 1488 C GLN B 61 25.511 -6.086 25.519 ATOM 1489 O GLN B 61 25.511 -6.086 25.519 ATOM 1490 CB GLN B 61 25.511 -6.086 25.519 ATOM 1491 CG GLN B 61 25.411 -4.866 24.746 ATOM 1492 CD GLN B 61 26.269 -6.763 26.028 ATOM 1499 CB GLN B 61 26.269 -6.763 26.028 ATOM 1499 CB GLN B 61 26.560 -4.382 24.774 ATOM 1499 CB GLN B 61 26.560 -4.382 24.774 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.560 -7.675 29.992 ATOM 1499 CB GLN B 61 26.600 -7.675 29.992 ATOM 1490 CB GLN B 61 26.600 -7.675 29.992 ATOM 1490 CB GLN B 61 26.600 -7.675 29.992 ATOM 1490 CB GLN B 61 26.600 -7.675 29.992 ATOM 1490 CB GLN B 61 26.								21.	788		
ATOM 1465 C TYR B 59 23.244 -8.290 20.607 ATOM 1466 O TYR B 59 22.178 -7.728 20.927 ATOM 1466 CB TYR B 59 24.387 -7.241 18.653 ATOM 1468 CG TYR B 59 24.271 -7.075 17.149 ATOM 1469 CD1 TYR B 59 24.271 -7.075 17.149 ATOM 1470 CD2 TYR B 59 25.385 -6.753 16.374 ATOM 1471 CE1 TYR B 59 25.385 -6.753 16.374 ATOM 1471 CE1 TYR B 59 22.939 -7.093 15.112 ATOM 1472 CE2 TYR B 59 25.291 -6.603 14.995 ATOM 1474 OH TYR B 59 24.068 -6.774 14.365 ATOM 1475 HH TYR B 59 24.068 -6.774 14.365 ATOM 1476 N ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 24.556 -7.595 23.615 ATOM 1481 CB ASP B 60 25.654 -7.261 23.125 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 22.803 -10.960 23.332 ATOM 1484 OD2 ASP B 60 22.803 -10.960 23.332 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 C GLN B 61 24.156 -7.022 24.774 ATOM 1487 CA GLN B 61 25.011 -6.086 25.513 ATOM 1488 C GLN B 61 25.511 -4.866 24.746 ATOM 1489 O GLN B 61 25.600 -4.382 24.832 ATOM 1489 C GLN B 61 25.714 -7.766 26.753 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.560 -4.382 24.832 ATOM 1493 OE1 GLN B 61 26.560 -7.675 29.992 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.560 -7.675 29.992 ATOM 1493 OE1 GLN B 61 26.560 -7.675 29.992 ATOM 1494 NE2 GLN B 61 26.560 -7.675 29.992 ATOM 1499 CD GLN B 61 26.560 -7.675 29.992 ATOM 1499 CD GLN B 61 26.574 -7.266 28.185 ATOM 1499 CD GLN B 61 26.560 -7.675 29.992 ATOM 1491 CG GLN B 61 26.560 -7.675 29.992 ATOM 1492 CD GLN B 61 26.571 -7.766 26.753 ATOM 1493 OE1 GLN B 61 26.560 -7.675 29.992 ATOM 1490 CB GLN B 61 26.560 -7.675 29.992 ATOM 1491 CG GLN B 61 26.560 -7.675 29.992 ATOM 1492 CD GLN B 61 26.560 -7.675 29.992 ATOM 1493 OE1 GLN B 61 26.560 -7.675 29.992 ATOM 1494 NE2 GLN B 61 26.560 -7.675 29.992 ATOM 1495 CB ILE B 62 24.571 -1.885 24.144 ATOM 1496 2HE2 GLN B 61 26.579 -4.579 23.933 ATOM 1499 CD ILE B						59		23.	631		
ATOM 1466 O TYR B 59						59		23.	244		
ATOM 1468 CG TYR B 59 24.387 -7.241 18.653 ATOM 1468 CG TYR B 59 24.271 -7.075 17.149 ATOM 1469 CD1 TYR B 59 23.045 -7.242 16.494 ATOM 1470 CD2 TYR B 59 25.385 -6.753 16.374 ATOM 1471 CE1 TYR B 59 25.385 -6.753 16.374 ATOM 1472 CE2 TYR B 59 25.291 -6.603 14.995 ATOM 1473 CZ TYR B 59 25.291 -6.603 14.995 ATOM 1474 OH TYR B 59 24.068 -6.774 14.365 ATOM 1475 HH TYR B 59 24.068 -6.774 14.365 ATOM 1476 N ASP B 60 24.010 -8.785 21.558 ATOM 1477 H ASP B 60 24.010 -8.785 21.598 ATOM 1477 H ASP B 60 24.010 -8.785 21.372 ATOM 1478 CA ASP B 60 24.010 -8.785 21.596 ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1481 CB ASP B 60 25.654 -7.261 23.125 ATOM 1482 CG ASP B 60 25.654 -7.261 23.125 ATOM 1484 OD2 ASP B 60 22.803 -10.960 23.373 ATOM 1484 OD2 ASP B 60 21.619 -10.634 23.032 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 24.156 -7.022 24.774 ATOM 1487 CA GLN B 61 24.156 -7.022 24.774 ATOM 1488 C GLN B 61 25.011 -6.086 25.513 ATOM 1489 O GLN B 61 25.011 -6.086 25.533 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.269 -6.763 26.028 ATOM 1493 OE1 GLN B 61 26.269 -6.763 26.028 ATOM 1494 NEC GLN B 61 26.269 -6.765 23.933 ATOM 1498 C GLN B 61 26.560 -4.382 24.832 ATOM 1499 CB GLN B 61 26.560 -7.865 28.185 ATOM 1499 CB GLN B 61 26.560 -7.865 28.185 ATOM 1499 CB GLN B 61 26.560 -7.865 29.993 ATOM 1499 CB GLN B 61 26.560 -7.865 29.993 ATOM 1499 CB GLN B 61 26.560 -7.864 29.014 ATOM 1498 H ILE B 62 24.571 -7.846 23.233 ATOM 1499 CB GLN B 61 26.560 -7.675 29.993 ATOM 1499 CB GLN B 61 26.560 -7.664 23.233 ATOM 1499 CB GLN B 61 26.560 -7.665 28.185 ATOM 1499 CB GLN B 61 26.560 -7.675 29.993 ATOM 1490 CB GLN B 61 26.560 -7.665 28.185 ATOM 1490 CB GLN B 61 26.560 -7.665 28.185 ATOM 1490 CB GLN B 61 26.560 -7.675 29.993 ATOM 1490 CB GLN B 61 26.560 -7.675 29.993 ATOM 1490 CB GLN B 61 26.560 -7.675 29.993 ATOM 1490 CB GLN B 61 26.560 -7.675 29.993 ATOM 1490 CB GLN B 61 26.560 -7.654 28.185 ATOM 1490 CB GLN B 61 26.560 -7.654 28.185 ATOM 1490 CB GLN B 61 26.						59					
ATOM 1468 CG TYR B 59 24.271 -7.075 17.149 ATOM 1469 CD1 TYR B 59 23.045 -7.242 16.494 ATOM 1470 CD2 TYR B 59 25.385 -6.753 16.374 ATOM 1471 CE1 TYR B 59 25.385 -6.753 16.374 ATOM 1472 CE2 TYR B 59 22.939 -7.093 15.112 ATOM 1473 CZ TYR B 59 22.939 -7.093 15.112 ATOM 1474 CH TYR B 59 24.068 -6.774 14.365 ATOM 1475 HH TYR B 59 24.068 -6.774 14.365 ATOM 1476 N ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 24.556 -7.595 23.615 ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 24.556 -7.595 23.615 ATOM 1481 CB ASP B 60 22.803 -10.960 23.332 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 21.619 -10.634 23.032 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 24.156 -7.022 24.774 ATOM 1487 CA GLN B 61 23.252 -7.234 25.146 ATOM 1488 C GLN B 61 25.511 -6.086 25.519 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1489 O GLN B 61 25.411 -4.866 24.746 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.560 -4.382 24.832 ATOM 1493 OE1 GLN B 61 26.560 -4.382 24.832 ATOM 1499 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.571 -7.66 28.188 ATOM 1493 OE1 GLN B 61 26.560 -4.382 24.832 ATOM 1494 NE2 GLN B 61 26.571 -7.66 28.188 ATOM 1495 LHE2 GLN B 61 26.571 -7.66 28.588 ATOM 1496 AHE2 GLN B 61 26.571 -7.844 29.014 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.571 -1.885 24.144 ATOM 1496 AHE2 GLN B 61 26.568 -1.709 21.094 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.579 -4.279 20.878 ATOM 1500 C ILE B 62 24.500 1.296 22.5015				TYR	В	59					
ATOM 1470 CD2 TYR B 59 23.045 -7.242 16.494 ATOM 1471 CD2 TYR B 59 25.385 -6.753 16.374 ATOM 1471 CE1 TYR B 59 22.939 -7.093 15.112 ATOM 1472 CE2 TYR B 59 22.939 -7.093 15.112 ATOM 1473 CZ TYR B 59 24.068 -6.774 14.365 ATOM 1474 OH TYR B 59 24.018 -6.620 13.010 ATOM 1475 HH TYR B 59 24.018 -6.620 13.010 ATOM 1476 N ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.010 -8.785 21.596 ATOM 1477 CA ASP B 60 24.852 -9.276 21.372 ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 24.556 -7.595 23.615 ATOM 1481 CB ASP B 60 23.789 -9.920 23.777 ATOM 1482 CG ASP B 60 23.789 -9.920 23.777 ATOM 1483 OD1 ASP B 60 23.789 -9.920 23.777 ATOM 1484 OD2 ASP B 60 23.208 -10.960 23.32 ATOM 1485 N GLN B 61 23.252 -7.234 25.146 ATOM 1486 H GLN B 61 24.156 -7.022 24.774 ATOM 1486 C GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -6.086 25.519 ATOM 1489 O GLN B 61 26.269 -6.763 26.028 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.269 -6.763 26.028 ATOM 1492 CD GLN B 61 26.269 -6.763 26.028 ATOM 1494 NE2 GLN B 61 26.269 -6.763 26.028 ATOM 1494 NE2 GLN B 61 26.269 -6.763 28.185 ATOM 1495 1HE2 GLN B 61 26.269 -6.763 28.185 ATOM 1496 H LE B 62 24.572 -7.455 28.548 ATOM 1497 N ILE B 62 24.572 -7.455 28.548 ATOM 1498 H LLE B 62 24.571 -1.885 24.144 ATOM 1498 H LLE B 62 24.571 -1.885 24.144 ATOM 1498 H LLE B 62 24.571 -1.885 24.144 ATOM 1498 H LLE B 62 24.571 -1.885 24.144 ATOM 1498 H LLE B 62 24.571 -1.885 24.144 ATOM 1498 H LLE B 62 24.571 -1.885 24.144 ATOM 1498 H LLE B 62 24.571 -1.885 24.144 ATOM 1500 CB ILE B 62 24.571 -1.885 24.144 ATOM 1500 CB ILE B 62 24.571 -1.885 24.144 ATOM 1500 CB ILE B 62 24.571 -1.885 24.144 ATOM 1500 CB ILE B 62 24.571 -1.885 24.144 ATOM 1500 CB ILE B 62 24.572 -7.292 21.912 ATOM 1500 CB ILE B 62 24.573 -4.479 20.878 ATOM 1500 CB ILE B 62 24.571 -1.028 23.238 ATOM 1500 CB ILE B 62 24.568 -0.912 24.304 ATOM 1500 CB ILE B 62 24.568 -0.912 24.304 ATOM 1500 CB ILE B 62 24.560 -1.709 20.878 ATOM 1500 CD1 ILE B 62 24.560 -1.709 20.878				TYR	В	59					
ATOM 1470 CD2 TYR B 59 25.385 -6.753 16.374 ATOM 1471 CE1 TYR B 59 22.939 -7.093 15.112 ATOM 1472 CE2 TYR B 59 25.291 -6.603 14.995 ATOM 1473 CZ TYR B 59 24.068 -6.774 14.365 ATOM 1474 OH TYR B 59 24.018 -6.620 13.010 ATOM 1475 HH TYR B 59 24.018 -6.620 13.010 ATOM 1475 HH TYR B 59 24.926 -6.394 12.658 ATOM 1476 N ASP B 60 24.926 -6.394 12.658 ATOM 1477 H ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 23.644 -8.624 22.992 ATOM 1479 C ASP B 60 23.644 -8.624 22.992 ATOM 1480 O ASP B 60 23.789 -9.920 23.777 ATOM 1481 CB ASP B 60 23.789 -9.920 23.777 ATOM 1482 CG ASP B 60 23.789 -9.920 23.777 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1485 N GLN B 61 24.156 -7.021 24.774 ATOM 1486 H GLN B 61 24.156 -7.022 24.774 ATOM 1487 CA GLN B 61 24.156 -7.022 24.774 ATOM 1488 C GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -6.086 25.519 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.560 -4.382 24.832 ATOM 1494 NE2 GLN B 61 26.560 -4.382 24.832 ATOM 1495 IHE2 GLN B 61 26.571 -7.765 28.156 ATOM 1497 N ILE B 62 24.571 -7.844 29.014 ATOM 1498 H ILE B 62 24.571 -1.885 24.144 ATOM 1498 H ILE B 62 24.571 -1.885 24.184 ATOM 1499 CA ILE B 62 24.571 -1.885 24.184 ATOM 1499 CA ILE B 62 24.571 -1.885 24.184 ATOM 1499 CA ILE B 62 24.571 -1.885 24.184 ATOM 1490 CB GLN B 61 26.568 -4.572 23.933 ATOM 1490 CC ILE B 62 24.571 -1.885 24.144 ATOM 1498 H ILE B 62 24.571 -1.885 24.144 ATOM 1498 H ILE B 62 24.571 -1.885 24.144 ATOM 1498 H ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 63 25.585 -0.912 24.304				TYR	В						
ATOM 1471 CE1 TYR B 59			CD2	TYR	В	59					
ATOM 1473 CZ TYR B 59		1471	CEl	TYR	В						
ATOM 1474 OH TYR B 59		1472	CE2	TYR	В						
ATOM 1475 HH TYR B 59 24.926 -6.394 12.658 ATOM 1476 N ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 24.852 -9.276 21.372 ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 25.654 -7.261 23.125 ATOM 1483 OD1 ASP B 60 23.789 -9.920 23.777 ATOM 1484 OD2 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1486 H GLN B 61 25.011 -6.086 25.519 ATOM 1489 O GLN B 61 25.411 -4.866 24.746 ATOM 1489 O GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.269 -6.763 26.028 ATOM 1493 OE1 GLN B 61 26.5714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.5744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.5744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.560 -7.675 28.548 ATOM 1497 N ILE B 62 24.879 -4.257 23.933 ATOM 1498 H ILE B 62 24.879 -4.257 23.933 ATOM 1498 H ILE B 62 24.878 -3.047 23.238 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.879 -2.922 21.912 ATOM 1501 CG ILE B 62 24.879 -4.257 23.933 ATOM 1500 C ILE B 62 24.879 -4.257 23.933 ATOM 1500 C ILE B 62 24.879 -4.257 23.933 ATOM 1500 C ILE B 62 24.879 -2.922 21.912 ATOM 1501 CG ILE B 62 24.879 -2.922 21.912 ATOM 1505 CD ILE B 62 24.571 -1.885 24.144 ATOM 1501 C ILE B 62 24.879 -2.922 21.912 ATOM 1505 CD ILE B 62 24.570 -4.479 20.878 ATOM 1506 N LEU B 63 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.300 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.		1473	ĊZ	TYR	В						
ATOM 1476 N ASP B 60 24.010 -8.785 21.596 ATOM 1477 H ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 23.644 -8.624 22.992 ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 25.654 -7.261 23.125 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 21.619 -10.634 23.032 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.411 -4.866 24.746 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.560 -4.382 24.832 ATOM 1493 OEI GLN B 61 26.560 -7.326.028 ATOM 1494 NE2 GLN B 61 26.5714 -7.766 28.185 ATOM 1495 1HE2 GLN B 61 26.5714 -7.844 29.014 ATOM 1496 2HE2 GLN B 61 26.620 -7.675 28.548 ATOM 1499 CA ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.510 -1.819 24.819 ATOM 1500 C ILE B 62 24.510 -4.170 21.094 ATOM 1501 C ILE B 62 24.510 -4.170 21.094 ATOM 1505 CD1 ILE B 62 24.510 -4.170 21.094 ATOM 1505 CD1 ILE B 62 24.510 -4.170 21.094 ATOM 1506 N LEU B 63 25.492 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015	MOTA	1474	OH	TYR	В						
ATOM 1477 H ASP B 60 24.852 -9.276 21.372 ATOM 1478 CA ASP B 60 23.644 -8.624 22.992 ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 23.789 -9.920 23.777 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 21.619 -10.634 23.032 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -6.086 24.746 ATOM 1489 O GLN B 61 25.011 -6.086 25.519 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.560 -8.038 26.753 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 HEE2 GLN B 61 26.744 -7.844 29.014 ATOM 1496 2HE2 GLN B 61 26.744 -7.844 29.014 ATOM 1497 N ILE B 62 24.572 -7.455 28.548 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1500 CGI ILE B 62 24.5768 -1.709 21.067 ATOM 1506 N LEU B 63 25.794 -4.479 20.878 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 25.592 0.322 25.015		1475	HH								
ATOM 1478 CA ASP B 60 23.644 -8.624 22.992 ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 23.789 -9.920 23.777 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 24.156 -7.022 24.774 ATOM 1486 C GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -6.086 25.519 ATOM 1489 O GLN B 61 25.411 -4.866 24.746 ATOM 1489 C GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.560 -4.382 24.832 ATOM 1492 CD GLN B 61 26.560 -7.765 28.185 ATOM 1493 OE1 GLN B 61 26.744 -7.766 28.185 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.571 -1.885 24.144 ATOM 1499 CA ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.571 -1.885 24.144 ATOM 1504 CG2 ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.568 -1.709 21.067 ATOM 1500 C ILE B 62 24.568 -1.709 21.067 ATOM 1500 C ILE B 62 24.568 -1.709 21.067 ATOM 1500 C ILE B 63 25.192 0.322 25.015 ATOM 1500 C LEU B 63 25.192 0.322 25.015	MOTA	1476	N								
ATOM 1479 C ASP B 60 24.556 -7.595 23.615 ATOM 1480 O ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 23.789 -9.920 23.777 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 N GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -4.866 24.746 ATOM 1489 O GLN B 61 25.411 -4.866 24.746 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.269 -6.763 26.028 ATOM 1492 CD GLN B 61 26.020 -8.038 26.753 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.571 -1.885 24.144 ATOM 1504 CG2 ILE B 62 24.571 -1.885 24.144 ATOM 1505 CD1 ILE B 62 24.568 -1.709 21.067 ATOM 1506 N LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015	ATOM	1477									
ATOM 1480 O ASP B 60 25.654 -7.261 23.125 ATOM 1481 CB ASP B 60 23.789 -9.920 23.777 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -4.866 24.746 ATOM 1489 O GLN B 61 25.411 -4.866 24.746 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.269 -6.763 26.028 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 25.714 -7.765 28.185 ATOM 1494 NE2 GLN B 61 24.572 -7.455 28.548 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 26.564 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1503 CGI ILE B 62 24.571 -1.885 24.144 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CDI ILE B 62 24.568 -1.709 21.067 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.492 -0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 26.403 1.296 24.030	MOTA	1478									
ATOM 1481 CB ASP B 60 23.789 -9.920 23.777 ATOM 1482 CG ASP B 60 22.803 -10.960 23.332 ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -6.086 25.519 ATOM 1489 O GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.269 -6.762 26.028 ATOM 1492 CD GLN B 61 26.020 -8.038 26.753 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.720 -7.455 28.548 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.571 -1.885 24.144 ATOM 1504 CG2 ILE B 62 24.579 -2.922 21.912 ATOM 1505 CD1 ILE B 62 24.579 -4.479 20.878 ATOM 1506 N LEU B 63 25.794 -4.479 20.878 ATOM 1507 H LEU B 63 25.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 24.630 1.296 24.030	ATOM	1479	С								
ATOM 1482 CG ASP B 60	MOTA	1480									
ATOM 1483 OD1 ASP B 60 21.619 -10.634 23.032 ATOM 1484 OD2 ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.411 -4.866 24.746 ATOM 1489 O GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.269 -6.763 26.028 ATOM 1492 CD GLN B 61 26.020 -8.038 26.753 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 24.572 -7.455 28.548 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1498 CA ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.568 -1.709 21.067 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.094 ATOM 1505 CD1 ILE B 62 24.568 -1.709 21.094 ATOM 1506 N LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.680 1.296 24.030	MOTA	1481									
ATOM 1484 OD2 ASP B 60 23.208 -12.126 23.273 ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.411 -4.866 24.746 ATOM 1489 O GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.744 -7.844 29.014 ATOM 1496 2HE2 GLN B 61 26.744 -7.844 29.014 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.576 -1.819 24.819 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 24.568 -1.709 21.067 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.493 -1.296 24.304 ATOM 1508 CA LEU B 63 25.493 0.322 25.015 ATOM 1508 CA LEU B 63 25.493 0.222 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030	MOTA	1482									
ATOM 1485 N GLN B 61 24.156 -7.022 24.774 ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.011 -6.086 24.746 ATOM 1489 O GLN B 61 25.411 -4.866 24.746 ATOM 1490 CB GLN B 61 26.560 -4.382 24.832 ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 26.020 -8.038 26.753 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.744 -7.844 29.014 ATOM 1496 2HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.571 -1.885 24.144 ATOM 1504 CG2 ILE B 62 24.570 -2.922 21.912 ATOM 1505 CD1 ILE B 62 24.579 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.495 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.690 1.296 24.030	MOTA										
ATOM 1486 H GLN B 61 23.252 -7.234 25.146 ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.411 -4.866 24.746 ATOM 1489 O GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.572 -7.455 28.548 ATOM 1494 NE2 GLN B 61 26.620 -7.675 29.992 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.568 -1.709 21.067 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 24.568 -1.709 21.067 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.493 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.600 1.296 24.030											
ATOM 1487 CA GLN B 61 25.011 -6.086 25.519 ATOM 1488 C GLN B 61 25.411 -4.866 24.746 ATOM 1489 O GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 24.572 -7.455 28.548 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.571 -1.885 24.144 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 24.630 1.296 24.036											
ATOM 1488 C GLN B 61 25.411 -4.866 24.746 ATOM 1489 O GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 24.572 -7.455 28.548 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.571 -1.885 24.144 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 24.568 -1.709 21.067 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 24.630 1.296 24.030											
ATOM 1488 C GLN B 61 26.560 -4.382 24.832 ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 25.714 -7.766 28.185 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 24.539 -4.257 23.933 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.571 -1.885 24.144 ATOM 1503 CG1 ILE B 62 24.571 -1.819 24.819 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 24.568 -1.709 21.067 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 24.630 1.296 24.030											
ATOM 1490 CB GLN B 61 26.269 -6.763 26.028 ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 24.572 -7.455 28.548 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030											
ATOM 1491 CG GLN B 61 26.020 -8.038 26.753 ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 24.572 -7.455 28.548 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 23.515 -1.819 24.819 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 22.630 1.296 24.030											
ATOM 1492 CD GLN B 61 25.714 -7.766 28.185 ATOM 1493 OE1 GLN B 61 24.572 -7.455 28.548 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 24.571 -1.885 24.144 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 24.630 1.296 24.030											
ATOM 1493 OE1 GLN B 61 24.572 -7.455 28.548 ATOM 1494 NE2 GLN B 61 26.744 -7.844 29.014 ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1508 CA LEU B 63 24.630 1.296 24.030											
ATOM 1494 NE2 GLN B 61 26.620 -7.675 29.992 ATOM 1495 1HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030											
ATOM 1495 1HE2 GLN B 61 26.620 -7.675 29.992 ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030											29.014
ATOM 1496 2HE2 GLN B 61 27.654 -8.073 28.669 ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030											29.992
ATOM 1497 N ILE B 62 24.539 -4.257 23.933 ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030										-8.073	28.669
ATOM 1498 H ILE B 62 23.628 -4.648 23.801 ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 25.485 -0.912 24.304 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030										-4.257	
ATOM 1499 CA ILE B 62 24.878 -3.047 23.238 ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030										-4.648	
ATOM 1500 C ILE B 62 24.571 -1.885 24.144 ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030										-3.047	
ATOM 1501 O ILE B 62 23.515 -1.819 24.819 ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030										-1.885	
ATOM 1502 CB ILE B 62 24.097 -2.922 21.912 ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030										-1.819	
ATOM 1503 CG1 ILE B 62 24.310 -4.170 21.094 ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030			_							-2.922	
ATOM 1504 CG2 ILE B 62 24.568 -1.709 21.067 ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030											
ATOM 1505 CD1 ILE B 62 25.794 -4.479 20.878 ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030				_							
ATOM 1506 N LEU B 63 25.485 -0.912 24.304 ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030								25.	.794		
ATOM 1507 H LEU B 63 26.403 -1.028 23.926 ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030											
ATOM 1508 CA LEU B 63 25.192 0.322 25.015 ATOM 1509 C LEU B 63 24.630 1.296 24.030											
ATOM 1509 C LEU B 63 24.630 1.296 24.030								25	.192		
05 000 1 (50 00 00						63					
						63		25	.239	1.658	22.995

FIG. I laa

			41	/46			
ATOM	1511	СВ	LEU B	63	26.436	0.970	25.590 26.226
ATOM	1512		LEU B	63	26.186	2.358	27.576
MOTA	1513		LEU B	63	25.486 27.468	3.162	26.382
ATOM	1514		LEU B	63	23.492	1.946	24.358
MOTA	1515	N	ILE B	64	22.958	1.643	25.148
MOTA	1516	H	ILE B	64 64	23.003	3.068	23.617
ATOM	1517	CA	ILE B	64	22.872	4.194	24.612
MOTA	1518	C 0	ILE B	64	22.915	4.007	25.846
ATOM	1519 1520	СВ	ILE B	64	21.634	2.701	22.989
MOTA MOTA	1521	CG1	ILE B	64	21.825	1.521	22.029
ATOM	1522	CG2	ILE B	64	20.982	3.894	22.246
ATOM	1523	CD1	ILE B	64	20.593	1.096	21.260
ATOM	1524	N	GLU B	65	22.803	5.460	24.172 23.216
ATOM	1525	H	GLU B	65	23.013	5.664 6.551	25.037
MOTA	1526	CA	GLU B	65	22.432 21.242	7.194	24.373
MOTA	1527	C	GLU B	65	21.312	7.729	23.257
MOTA	1528	0	GLU B	65 65	23.497	7.615	25.131
MOTA	1529	CB ·	GLU B	65	24.787	7.196	25.761
MOTA	1530	CG CD	GLU B	65	25.694	8.385	26.076
ATOM	1531 1532	OE1	GLU B	65	25.170	9.510	26.311
MOTA MOTA	1532	OE2	GLU B	65	26.938	8.200	26.092
ATOM	1534	N	ILE B	66	20.078	7.240	25.035
ATOM	1535	Н	ILE B	66	20.010	6.835	25.947
ATOM	1536	CA	ILE B	66	18.907	7.865	24.462
MOTA	1537	С	ILE B	66	18.777	9.195 9.303	25.145 26.379
MOTA	1538	0	ILE B	66	18.591	6.995	24.790
MOTA	1539	CB	ILE B	66	17.713 17.916	5.583	24.335
MOTA	1540	CG1	ILE B	66 66	16.405	7.544	24.177
MOTA	1541	CG2	ILE B ILE B	66	16.888	4.677	24.884
MOTA	1542 1543	CD1	CYS B	67	18.965	10.325	24.437
MOTA MOTA	1544	H	CYS B	67	19.201	10.268	23.467
ATOM	1545	CA	CYS B	67	18.833	11.663	25.049
ATOM	1546	C	CYS B	67	19.637	11.781	26.319
ATOM	1547	0	CYS. B	67	19.235	12.400	27.328 25.319
MOTA	1548	CB	CYS B	67	17.387	12.023	23.821
ATOM	1549	SG	CYS B	67	16.407 20.830	12.259 11.180	26.383
MOTA	1550	N	GLY B	68	21.158	10.646	25.604
MOTA	1551	H	GLY B	68 68	21.654	11.288	27.558
ATOM	1552	CA	GLY B	68	21.464	10.185	28.584
ATOM	1553	C O	GLY B	68	22.174	10.128	29.606
MOTA MOTA	1554 1555	Ŋ	HIS B	69	20.513	9.255	28.425
ATOM	1556	H	HIS B	69	19.924	9.282	27.618
ATOM	1557	CA	HIS B	69	20.304	8.199	29.391
ATOM	1558	C	HIS B	69	20.861	6.936	28.811
MOTA	1559	0	HIS B	69	20.589	6.560	27.647 29.654
ATOM	1560	CB	HIS B	69	18.832	7.992 9.203	30.223
ATOM	1561	CG	HIS B	69	18.175	9.203	31.435
ATOM	1562	MDI		69	17.504 17.383	8.402	32.032
ATOM	1563	HD1		69 69	18.122	10.470	29.729
MOTA	1564	CD2		69	17.070	10.429	31.626
MOTA	1565			69	17.410	11.240	30.635
MOTA	1566	NE2	HTD D	33			

FIG. 1 lbb

			42	146			
> mov4	1567	N	LYS B	70	21.751	6.217	29.499
ATOM	1567 1568	-	LYS B	70	22.025	6.512	30.414
ATOM			LYS B	70	22.326	5.020	28.945
MOTA	1569		LYS B	70	21.386	3.854	29.145
MOTA	1570		LYS B	70	20.627	3.725	30.120
MOTA	1571	_	LYS B	70	23.613	4.678	29.663
MOTA	1572		LYS B	70	24.694	5.655·	29.379
ATOM	1573	CG	LYS B	70	25.739	5.524	30.444
MOTA	1574	CD	LYS B	70	27.048	6.090	30.011
MOTA	1575	CE	LYS B	70	26.948	7.548	30.000
MOTA	1576	NZ	LYS B	70	27.821	7.940	29.711
MOTA		1HZ	LYS B	70	26.725	7.874	30.919
MOTA		3HZ	LYS B	70	26.230	7.828	29.363
MOTA		2HZ	ALA B	71	21.512	2.849	28.284
MOTA	1580	N	ALA B	71	22.141	2.934	27.512
MOTA	1581	Н	ALA B	71	20.762	1.630	28.432
MOTA	1582	CA		71	21.629	0.576	27.805
MOTA	1583	C		71	22.463	0.830	26.912
MOTA	1584	0_	ALA B	71	19.452	1.726	27.737
MOTA	1585	CB	ALA B	72	21.547	-0.681	28.237
MOTA	1586	N	ILE B	72	20.864	-0.925	28.926
MOTA	1587	H	ILE B	72	22.424	-1.698	27.730
MOTA	1588	CA	ILE B	72	21.615	-2.938	27.462
MOTA	1589	C	ILE B	72 72	20.909	-3.490	28.330
MOTA	1590	0	ILE B	72	23.524	-1.999	28.737
MOTA	1591	CB	ILE B	72	24.322	-0.735	29.090
MOTA	1592	CG1	ILE B	72	24.442	-3.037	28.153
MOTA	1593	CG2	ILE B	72	25.374	-1.012	30.163
MOTA	1594	CD1	ILE B	73	21.609	-3.446	26.235
MOTA	1595	N	GLY B	73 73	22.204	-3.054	25.534
MOTA	1596	H	GLY B	73 73	20.707	-4.545	26.062
MOTA	1597	CA	GLY B	73 73	20.828	-5.084	24.663
MOTA	1598	C		73 73	21.754	-4.831	23.863
MOTA	1599	0		74	19.856	-5.905	24.271
MOTA	1600	N	-	74	19.086	-6.088	24.882
MOTA	1601	H	THR B	74	19.869	-6.548	22.988
MOTA	1602	CA		74	19.363	-5.590	21.931
MOTA	1603	C	THR B	74 74	18.338	-4.870	22.053
MOTA	1604	0	THR B	74	19.011	-7.801	23.074
MOTA	1605			74	19.611	-8.683	24.013
MOTA	1606			74	19.068	-9.519	24.092
MOTA	1607			74	18.817	-8.496	21.705
MOTA	1608			75	20.028	-5.620	20.762
MOTA	1609		VAL B	75	20.835	-6.203	20.666
ATOM	1610			75 75	19.630	-4.837	19.611
ATOM	1611		• •	75 75	19.600	-5.771	18.426
MOTA	1612			75 75	20.444	-6.673	18.230
MOTA	1613		VAL B	75 75	20.667	-3.712	19.395
ATOM	1614			75 75	20.473	-3.002	18.046
ATOM	1615		_	75	20.679	-2.708	20.567
MOTA	1616				18.557	-5.647	17.565
ATOM	1617				17.822	-5.000	17.767
ATOM	1618		LEU B		18.444	-6.427	16.324
MOTA	1619		LEU B		18.736	-5.487	15.144
MOTA	1620		LEU B		18.239	-4.343	15.040
MOTA	1621		LEU B		17.028	-7.021	16.158
MOTA	1622	2 CB	ם הפת				

FIG. I lcc

43/46

17.449 -7.612 16.427 76 LEU B CG 1623 MOTA 17.263 -8.075 14.992 76 CD1 LEU B 1624 MOTA 18.019 -8.758 17.266 CD2 LEU B 76 **ATOM** 1625 14.222 19.607 -5.900 VAL B 77 N MOTA 1626 14.276 -6.824 19.985 VAL B 77 1627 Н MOTA -5.042 13.133 20.027 77 VAL B CA MOTA 1628 11.842 -5.662 19.570 77 VAL B С MOTA 1629 -6.883 11.598 19.678 77 VAL B 0 1630 ATOM -4.905 13.191 21.563 77 VAL B CB MOTA 1631 -4.202 11.944 22.129 CG1 VAL B 77 MOTA 1632 14.470 -4.166 22.030 CG2 VAL B 77 1633 ATOM -4.915 10.943 18.978 GLY B 78 1634 N ATOM -3.941 11.121 18.841 GLY B 78 MOTA 1635 Н 9.705 -5.475 18.523 GLY B 78 ÇA MOTA 1636 8.874 -4.338 18.019 78 GLY B **ATOM** 1637 C 9.223 -3.14218.130 GLY B 78 0 1638 MOTA 7.722 -4.596 17.408 79 PRO B N 1639 MOTA -3.535 6.834 16.954 79 PRO B CA ATOM 1640 7.280 -2.87215.635 PRO B 79 C 1641 ATOM 6.565 -2.877 14.609 PRO B 79 0 1642 MOTA -4.274 5.492 16.804 79 PRO B CB MOTA 1643 16.463 -5.712 5.881 79 PRO B CG MOTA 1644 7.189 -5.959 17.159 PRO B 79 MOTA 1645 CD 8.458 -2.24715.574 80 THR B 1646 N MOTA -2.242 9.058 16.374 80 THR B Н 1647 MOTA -1.583 8.865 14.364 THR B 80 CA **ATOM** 1648 8.228 14.312 -0.189THR B 80 C MOTA 1649 0.471 8.001 15.349 THR B 80 0 1650 MOTA -1.512 10.410 14.250 80 THR B CB 1651 MOTA 10.806 13.079 -0.802 OG1 THR B 80 MOTA 1652 13.022 -0.766 11.804 80 THR B HG1 MOTA 1653 -0.901 11.062 15.519 80 THR B 1654 CG2 MOTA 0.354 7.885 13.137 PRO B 81 1655 N MOTA 7.379 13.036 1.747 PRO B 81 CA ATOM 1656 8.484 2.732 13.363 PRO B 81 C 1657 MOTA 8.250 13.791 3.880 PRO B 81 1658 0 MOTA 6.982 1.912 11.548 PRO B 81 1659 CB MOTA 7.488 0.674 10.819 CG PRO B 81 1660 MOTA 7.797 -0.387 11.854 CD PRO B 81 MOTA 1661 9.772 2.368 13.197 VAL B 82 MOTA 1662 N 9.992 1.427 12.940 VAL B 82 1663 Н MOTA 10.885 3.306 13.380 CA VAL B 82 1664 MOTA 12.010 14.160 2.668 C VAL B 82 1665 MOTA 12.293 1.465 14.045 VÁL B 82 1666 0 MOTA 11.431 3.695 11.996 VAL B 82 CB 1667 MOTA 12.269 4.961 12.055 82 1668 CG1 VAL B MOTA 10.318 10.958 3.857 CG2 VAL B 82 1669 MOTA 12.775 14.963 3.422 83 ASN B 1670 N MOTA 12.516 4.370 15.147 83 ASN B 1671 H MOTA 13.967 2.846 15.550 83 CA ASN B 1672 ATOM 15.022 2.874 14.481 ASN B 83 C MOTA 1673 15.294 3.903 13.814 83 ASN B 1674 0 MOTA 14.472 16.743 3.639 83 CB ASN B 1675 MOTA 13.570 17.935 3.574 83 1676 CG ASN B MOTA 13.167 18.409 2.511 OD1 ASN B 83 1677 MOTA 4.735 13.238 18.439 1678 ND2 ASN B MOTA

FIG. 1 Idd SUBSTITUTE SHEET (RULE 26)

44/46

				44	140			•
3 most	1670	2HD2	ASN I	3	83	19.237	4.786	12.638
ATOM			ASN I		83	18.030	5.580	13.582
MOTA			ILE E		84	14.225	1.749	15.711
MOTA	1681	-	ILE !		84	14.791	0.938	15.564
MOTA	1682		ILE I		84	13.154	1.658	16.667
ATOM	1683	-	ILE I		84	13.740	1.317	18.020
ATOM	1684			3 B	84	14.428	0.300-	18.223
MOTA	1685	0		В	84	12.214	0.517	16.260
ATOM	1686	CB		В	84	11.656	0.759	14.849
MOTA	1687	CG1		В	84	11.128	0.247	17.315
MOTA	1688	CG2		В	84	10.770	-0.359	14.291
MOTA	1689	CD1		В	85	13.483	2.157	19.051
ATOM	1690	N		В	85	13.028	3.030	18.877
ATOM	1691	H		В	85	13.846	1.834	20.408
MOTA	1692	CA			85	12.596	1.254	21.085
MOTA	1693	C	-	В	85	11.536	1.903	21.267
MOTA	1694	0		В	85	14.308	3.115	21.137
MOTA	1695	CB		B	85	15.447	3.826	20.395
ATOM	1696	CG1		В	85	14.673	2.840	22.589
MOTA	1697	CG2		В	85	16.730	3.053	20.263
MOTA	1698	CD1		В	86	12.617	-0.052	21.422
MOTA	1699	N		В		13.439	-0.595	21.251
MOTA	1700	H	GLY		86	11.481	-0.702	22.028
MOTA	1701	CA	_	В	86	11.557	-0.748	23.538
MOTA	1702	C	GLY		86	12.412	-0.165	24.238
MOTA	1703	0	GLY		86	10.614	-1.489	24.149
MOTA	1704	N	ARG		87	10.012	-2.072	23.604
MOTA	1705	H		В	87	10.442	-1.468	25.584
MOTA	1706	CA		В	87	11.627	-2.021	26.326
ATOM	1707	C	ARG		87	11.911	-1.666	27.495
MOTA	1708	0	-	В	87	9.200	-2.271	25.949
ATOM	1709	CB	ARG		87	7.951	-1.960	25.161
MOTA	1710	CG		В	87	6.956	-3.074	25.219
MOTA	1711	CD		В	87	5.906	-2.933	24.205
MOTA	1712	NE		В	87 87	5.790	-2.039	23.772
MOŢA	1713	HE		В		5.119	-3.953	23.856
MOTA	1714	CZ		В	87	5.252	-5.161	24.396
MOTA	1715	NH1		В	87	5.958	-5.326	25.085
MOTA	1716	2HH1		В	87	4.646	-5.905	24.113
MOTA	1717				87	4.180	-3.751	22.939
ATOM	1718	NH2			87	3.580	-4.502	22.664
ATOM	1719	1HH2		В	87	4.073	-2.848	22.524
MOTA	1720	2HH2	-	В	87	12.413	-2.937	25.731
MOTA	1721	N		В	88	12.206	-3.237	24.800
MOTA	1722	H		В	88	13.582	-3.519	26.415
MOTA	1723	CA	•	В	88	14.532	-2.429	26.821
MOTA	1724			В	88	15.214	-2.516	27.863
MOTA	1725		ASN	_	88	14.285	-4.605	25.559
ATOM	1726		ASN	В	88	15.063	-4.031	24.358
MOTA	1727		ASN	В	88	14.515	-3.245	23.612
atom	1728				88	16.333	-4.445	24.180
ATOM	1729			B	88	16.333	-4.099	23.414
MOTA	1730			B	88	16.744	-5.102	24.812
MOTA	1731				88	14.695	-1.328	26.061
MOTA	1732		LEU		89	14.695	-1.240	25.201
MOTA	1733		LEU		89	15.597	-0.234	26.452
ATOM	1734	CA	LEU	В	89	13.37	3.231	

FIG. | lee substitute sheet (RULE 26)

			45	5/46	•		
» TOM	1735	С	LEU B	89	14.797	0.937	27.053
MOTA	1736		LEU B	89	15.293	1.734	27.879
MOTA	1737		LEU B	89	16.421	0.232	25.236
MOTA	1738		LEU B	89		-0.754	24.567
MOTA	1739	_	LEU B	89	18.215	0.002	23.573
MOTA	1740		LEU B	89	18.352	-1.458	25.570
MOTA	1741	_	LEU B	90	13.511	1.114-	26.705
ATOM	1742		LEU B	90	13.082	0.486	26.056
ATOM ATOM	1743	CA	LEU B	90	12.698	2.221	27.257
ATOM	1744	C	LEU B	90	12.537	2.060	28.751
ATOM	1745	Ö	LEU B	90	12.575	3.033	29.533
ATOM	1746		LEU B	90	11.311	2.258	26.628
ATOM	1747		LEU B	90	11.232	2.730	25.168
ATOM	1748		LEU B	90	9.808	2.744	24.642
MOTA	1749		LEU B	90	11.831	4.105	24.982
ATOM	1750	N	THR B	91	12.315	0.843	29.271
ATOM	1751	Н	THR B	91	12.218 .	0.055	28.663
ATOM	1752	CA	THR B	91	12.210	0.634	30.699
ATOM	1753	С	THR B	91	13.537	1.028	31.375 32.518
ATOM	1754	0	THR B	91	13.575	1.525	
MOTA	1755	CB	THR B	91	=	-0.843	31.028 30.504
ATOM	1756	OG1	THR B	91		-1.676	30.504
ATOM	1757	HG1	THR B	91		-2.634	30.713
ATOM	1758	CG2	THR B	91		-1.285	30.418
ATOM	1759	N	GLN B	92	14.705	0.852	29.797
ATOM	1760	H	GLN B	92	14.707	0.497 1.190	31.433
MOTA	1761	CA	GLN B	92	15.920	2.660	31.633
MOTA	1762	С	GLN B	92	16.088	3.137	32.527
MOTA	1763	0	GLN B	92	16.807	0.680	30.682
MOTA	1764	CB	GLN B	92	17.127 17.076	-0.805	30.517
MOTA	1765	CG	GLN B	92	18.336	-1.314	29.900
ATOM	1766	CD	GLN B	92	19.394	-0.720	30.059
MOTA	1767	OE1	GLN B	92 92	18.221	-2.411	29.195
MOTA	1768	NE2	GLN B	92	19.022	-2.813	28.751
ATOM	1769	1HE2	GLN B	92 92	17.331	-2.856	29.095
ATOM	1770	2HE2	GLN B ILE B	93	15.538	3.512	30.746
MOTA	1771	N	ILE B	93	15.016	3.153	29.972
MOTA	1772	H		93	15.693	4.937	30.899
ATOM	1773	CA	ILE B	93	14.522	5.549	31.698
MOTA	1774	C 0	ILE B	93	14.438	6.773	31.940
ATOM	1775	CB	ILE B	93	15.981	5.657	29.548
MOTA	1776 1777	CG1		93	14.746	5.718	28.619
MOTA	1778	. CG2		93	17.223	5.060	28.874
MOTA	1779		_	93	14.946	6.734	27.488
ATOM	1780	N	GLY B	94	13.617	4.731	32.263
MOTA MOTA	1781	H	GLY B	94	13.639	3.752	32.060
MOTA	1782	CA	GLY B	94	12.594	5.224	33.170
MOTA	1783	C	GLY B	94	11.443	5.846	32.432
ATOM	1784	Õ	GLY B	94	10.766	6.803	32.878
atom	1785	Ŋ	CYS B	95	11.134	5.354	31.225
ATOM	1786	H	CYS B	95	11.603	4.538	30.888
ATOM	1787	CA	CYS B	95	10.134	5.969	30.381
ATOM	1788	C	CYS B	95	8.750	5.512	30.764
ATOM	1789		CYS B		8.478	4.309	31.006
ATOM	1790		CYS B	95	10.456	5.643	28.922

FIG. 1 Iff

ATOM	1791	SG	CYS	В	95	9.426	6.512	27.764
ATOM	1792	N	THR	В	96	7.778	6.444	30.764
ATOM	1793	H	THR	В	96	8.014	7.401	30.539
ATOM	1794	CA	THR	В	96	6.379	6.163	31.108
ATOM	1795	С	THR	В	96	5.390	6.970	30.254
ATOM	1796	0	THR	В	96	5.567	8.171	30.066
ATOM	1797	CB	THR	В	96	6.111	6.439	32.604
ATOM	1798	OG1	THR	В	96	6.341	7.794	32.938
ATOM	1799	HG1	THR	В	96	6.111	7.924	33.8 <i>6</i> 1
ATOM	1800	CG2	THR	В	96	6.938	5.566	33.554
ATOM	1801	N	LEU	B.	97	4.302	6.321	29.809
ATOM	1802	H	LEU	В	97	4.216	5.332	29.997
ATOM	1803	CA	LĒU	В	97	3.127	6.986	29.238
ATOM	1804	С	LEU	В	97	2.336	7.681	30.358
ATOM	1805	0	LEU	В	97	2.350	7221	31.499
ATOM	1806	CB	LEU	В	97	2.226	5.958	28.532
ATOM	1807	CG	LEU	В	97	2.860	5.279	27.300
ATOM	1808	CD1	LEU	В	97	2.101	3.986	26.957
ATOM	1809	CD2	LEU	В	97	2.842	6.216	26.085
ATOM	1810	N	ASN	В	98	1.637	8.777	30.024
ATOM	1811	Н	ASN	В	98	1.662	9.086	29.063
ATOM	1812	CA	ASN	В	98	0.906	9.631	30.960
ATOM	1813	С	ASN	В	98	-0.251	10.321	30.231
ATOM	1814	0	ASN	В	98	-0.032	11.303	29.522
ATOM	1815	CB	ASN	В	98	1.845	10.678	31.587
ATOM	1816	CG	ASN	В	98	2.783	10.077	32.634
ATOM	1817	OD1	ASN	В	98	3.926	9.739	32.335
ATOM	1818	ND2	ASN	В	98	2.297	9.942	33.870
ATOM	1819	2HD2	ASN	В	98	2.877	9.551	34.599
ATOM	1820	1HD2	ASN	В	98	1.351	10.229	34.074
ATOM	1821	N	LEU	В	99	-1.476	9.808	30.426
ATOM	1822	H	LEU	В	99	-1.568	9.010	31.037
ATOM	1823	CA	LEU	В	99	-2.709	10.288	29.797
ATOM	1824	С	LEU	В	99	-3.816	10.589	30.815
ATOM	1825	0	LEU	В	99	-3.630	10.272	32.011
ATOM	1826	CB	LEU	В	99	-3.146	9.340	28.657
ATOM	1827	CG	LEU	В	99	-3.714	7.932	28.941
MOTA	1828	CD1	LEU	В	99	-2.767	7.057	29.774
ATOM	1829	CD2	LEU	В	99	-5.134	7.943	29.528
MOTA	1830	OXT	LEU	В	99	-4.842	11.156	30.376
TER								

FIG. 1 1gg

<211> 6

-1-

SEQUENCE LISTING

```
<110> Kalyanaraman Ramnarayan
         Edward T. Maggio
         P. Patrick Hess
  <120> Use of Computationally Derived Protein Structures of Genetic Polymorphisms
in Pharmacogenomics for Drug Design and Clinical Applications
  <130> 24737-1906PC
  <140> Unassigned
  <141> 2000-11-10
  <150> 09/438,566
  <151> 1999-11-10
  <150> 24737-1906B
  <151> 2000-11-01
  <160> 118
  <170> FastSEQ for Windows Version 4.0
 <210> 1
 <211> 6
 <212> PRT
 <213> Artificial Sequence
 <220>
 <223> Modified Hepatitis C Virus NS3 Protease Inhibitor
       Peptide
 <221> ACETYLATION
 <222> 1
<221> MOD_RES
 <222> 2
<223> D-glutamic acid
<221> MOD_RES
<222> 5
<223> beta-cyclohexylalanine
<301> Ingallinella, P., Altamura, S., Bianchi, E., Talia <302> Potent Peptide Inhibitors Of Human Hepatitis C Vir
<303> Biochemistry
<304> 37
<305> 25
<306> 8906-8914
<307> 1998-06-23
<400> 1
Asp Xaa Leu Ile Xaa Cys
<210> 2
```

-2-

```
<212> PRT
 <213> Artificial Sequence
 <220>
 <223> Modified Hepatitis C Virus NS3 Protease Inhibitor
        Peptide
 <221> ACETYLATION
 <222> 1
 <221> MOD_RES
 <222> 5
 <223> beta-cyclohexylalanine
 <300>
<301> Ingallinella, P., Altamura, S., Bianchi, E., Talia <302> Potent Peptide Inhibitors Of Human Hepatitis C Vir
 <303> Biochemistry
 <304> 37
<305> 25
<306> 8906-8914
 <307> 1998-06-23
<400> 2
Asp Glu Leu Ile Xaa Cys
<210> 3
<211> 1045
<212> DNA
<213> Human Immunodificiency Virus (HIV)
<220>
<221> CDS
<222> (0)...(297)
<223> Protease
<221> CDS
<222> (298)...(1045)
<223> Portion of Reverse Transcriptase
cct cag atc act ctt tgg caa cga ccc cty gtc aca ata aag ata ggg
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly
                                                                                       48
                                             10
ggc caa cta aaa gaa gct yta tta gat aca gga gca gat gat aca gta
                                                                                      96
Gly Gln Leu Lys Glu Āla Xaa Leu Āsp Thr Gly Āla Āsp Āsp Thr Val
tta gaa gaa atg agt tta cca ggg aaa tgg aaa cca aaa atg ata ggg
                                                                                     144
Leu Glu Glu Met Ser Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
gga att gga ggt ttt atc aaa gta aga cag tat gat caa ata ctc ata
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Ile
                                                                                     192
gaa atc tgt gga cat aaa gct ata ggc aca gta tta gta gga cct aca
                                                                                     240
```

G.	lu I: 55	le C	ys G	ly H	is L	ys A] 70	la Il	e Gl	y Th	r Va 7	l Le 5	u Va	l Gl	y Pr	0 Thr		
Pi	et gt	c aa al Aa	ac a sn I	те т.	tt gg le G 35	ga ag ly Ar	ja aa rg As	t tt n Le	g ttg u Lei 90	ı Thi	t ca	g at n Il	t gg e Gl	у Су	c act s Thr	28	8
tt	a aa u As	it ti in Le	eu Pi	cc at ro Il 00	t ag Le Se	t co r Pr	t at	t gaa e Gli 10!	u Thi	gta Val	a cc	a gt o Va	a aa l Ly 11	s Le	a aag u Lys	33	6
Pr	a gg o Gl	a at y Me 11	E AS	at gg sp Gl	jc cc y Pr	a aa o Ly	a gti s Val 120	L Lys	a caa s Glr	tgg Trp	p cca	a ttg D Lei 12	u Th	a ga r Gl	a gaa u Glu	384	4
ц	13	0	SAI	а це	u va	1 G11 13!	u 116 5	e Cys	5 Thr	Glu	140	Gli	і Гу	s Gl:	a gga u Gly	432	2
14	5	- 56	т цу	SII	15	y Pro	o GIu	ı Asn	Pro	Tyr 155	Asn	1 Thr	Pro	o Ile	a ttt = Phe 160	480)
AT	a 116	я гуу	в гу	16	s As _] 5	o Sei	r Thr	Lys	170	Arg	Lys	Leu	Va]	Asr 175		528	i
VT.	, GI	T TIG	18	о гу	s Arç	g Thi	GIn	185	Phe	Trp	Glu	Val	Gln 190	Leu	a gga a Gly	576	
116	PIC	195	5 PIG	o Ala	a GIŽ	Leu	200	Gln	Lys	Lys	Ser	Val 205	Thr	Ile	ctg Leu	624	
vəř	210	GIY	/ ASI) Ala	туг	215	Ser	Val	Pro	Leu	Asp 220	Glu	Gly	Phe		. 672	
225	TYL	1111	Alc	Pne	230	TTE	cct Pro	Ser	Arg	Asn 235	Asn	Glu	Thr	Pro	Gly 240	720	
110	Arg	TYL	GII	245	ASII	val	ctc Leu	Pro	GIn 250	Gly	Trp	Lys	Gly	Ser 255	Pro	768	
AIG.	116	FILE	260	ser	ser	Met	aca Thr	Arg 265	Xaa	Leu	Glu	Pro	Phe 270	Arg	Lys	816	
G 211	Vell	275	GIU	116	AST	TIE	tat Tyr 280	GIn	Tyr 1	Met .	Asp	Asp 285	Leu	Tyr	Val	864	
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gga Gly 295	cag Gln	cat His	aga (Arg)	Ala:	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912	
aga	gga	cat	cta	tta	aag	tgg	gga	ttt a	acc a	aca d	cca	gac	aaa	aaa	cat	960	

-4-

Ar 30	g Gl 5	y Hi	s Le	u Le	u Ly 31	s Tr	p Gl	y Ph	e Thi	Th:		o As	p Ly	s Ly	s His 320	
Ca: Gl:	g aa n Ly	a ga s Gl	a cc u Pr	t cc o Pr 32	o Ph	t cti e Lei	tgg ıTr	g ato p Mei	g ggt t Gly 330	Ty	t gaa r Glu	a cto u Leo	c ca u Hi	t cc s Pro 33	t gat o Asp 5	1008
aaa Lys	a tg: s Tr]	g acap Thi	a gta r Vai 340	l Gl	g cct n Pro	t ata	a aag E Lys	g ttg 5 Let 345	ı Pro	gaa Glu	a aaa 1 Lys	ag s				1045
<21 <21	LO> 4 L1> 1 L2> I L3> I	1046 DNA	ı Imm	nuno	dific	cienc	y Vi	.rus	(HIV	')						
<22	21> C	(0)	.(29 Prote	7) ease												
<22		298)	(.on c		i) V Re	vers	e Tr	ansc	ript	ase						
cct	0> 4 cag Gln	ato	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	ctt Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile	gga Gly	48
Gly 999	cag Gln	cta Leu	aag Lys 20	Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gta Val	96
gtt Val	gaa Glu	gaa Glu 35	atg Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	aaa Lys	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	gly aaa	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gag Glu 60	caa Gln	ata Ile	gcc Ala	gta Val	192
gaa Glu 65	aty Xaa	tgt Cys	gga Gly	cat His	aga Arg 70	gct Ala	atg Met	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa	ata	aaa	gca	tta	gta	gaa	atc	tgt	aca	gaa	ttg	gaa	aag	gaa	9 99	432

Lys Ile L 130	ys Ala Leu Va	l Glu Ile Cys 135	s Thr Glu Leu Glu Lys Glu Gly 140	
145	15)	c cca tac aat act cca gta ttt n Pro Tyr Asn Thr Pro Val Phe 155 160	480
Ala Ile Ly	ag aaa aag aad ys Lys Asr 165	e agt act aaa Ser Thr Lys	tgg aga aaa tta gta gat ttc Trp Arg Lys Leu Val Asp Phe 170 175	528
aga gaa ct Arg Glu Le	t aat aag ags u Asn Lys Arg 180	act caa gac Thr Gln Asp 185	ttc tgg gag gtt caa tta gga Phe Trp Glu Val Gln Leu Gly 190	576
ata cca ca Ile Pro Hi 19	t cca gca ggg s Pro Ala Gly 5	tta aaa aag Leu Lys Lys 200	aat aaa tca ata aca gta ctg Asn Lys Ser Ile Thr Val Leu 205	624
gat gtg gg Asp Val Gl 210	t gat gca tat y Asp Ala Tyr	ttt tca gtt Phe Ser Val 215	ccc tta tgt gaa gac ttc agg Pro Leu Cys Glu Asp Phe Arg 220	672
aag tat act Lys Tyr Thi 225	t gca ttt acc r Ala Phe Thr 230	ata cct agt Ile Pro Ser	gta aac aat gag act cca ggg Val Asn Asn Glu Thr Pro Gly 235 240	720
·	245	var bea F10	cag gga tgg aaa gga ttc acc Gln Gly Trp Lys Gly Phe Thr 250 255	768
	260	265	atc tta gag cct ttt aga aaa Ile Leu Glu Pro Phe Arg Lys 270	816
caa aat cca Gln Asn Pro 275	gag ata gtt Glu Ile Val	atc tat caa : Ile Tyr Gln : 280	tac atg gat gat ttg tat gta Tyr Met Asp Asp Leu Tyr Val 285	864
gga tct gac Gly Ser Asp 290		ggg cag cat a Gly Gln His <i>P</i> 195	aga gca aaa ata gag gaa ctg arg Ala Lys Ile Glu Glu Leu 300	912
aga caa tat Arg Gln Tyr 305	ctg tgg aag t Leu Trp Lys 3	gg gga ttt t rp Gly Phe C	gc aca cca gaa caa aar cat ys Thr Pro Glu Gln Lys His 315 320	960
cag aaa gaa Gln Lys Glu	Pro Pro Phe I	er trb wer d	gt tat gaa ctc cat ccc gat ly Tyr Glu Leu His Pro Asp 30 335	1008
	gta caa cct a Val Gln Pro I 340	ta gtg ctg c le Val Leu P 345	ca gac aaa ga ro Asp Lys	1046

<210> 5
<211> 1104
<212> DNA
<213> Human Immunodificiency Virus (HIV)

-6-

<220> <221> CDS <222> (0)...(297) <223> HIV Protease <221> CDS <222> (298)...(1104) <223> Portion of HIV Reverse Transcriptase cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag rta ggg 48 Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Xaa Gly ggg caa cta agg gaa gct cta tta gat aca gga gca gat gat aca ata 96 Gly Gln Leu Arg Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Ile ata gaa gac ata act ttg cca gga aga tgg aca cca aaa atg ata ggg 144 Ile Glu Asp Ile Thr Leu Pro Gly Arg Trp Thr Pro Lys Met Ile Gly gga att gga ggt ttt gtc aaa gta aga cag tat gat cag ata ccc ata Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile 192 gaa atc tgc gga cat aaa rtt ata agt aca gta ttg gta gga cct aca Glu Ile Cys Gly His Lys Xaa Ile Ser Thr Val Leu Val Gly Pro Thr cca ata aac ata gtt gga aga aat ctg atg act cag att ggt tgc act Pro Ile Asn Ile Val Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr 288 tta aat ttt ccc att agt cct att gaa act gta cca gtc aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 336 105 cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 384 aaa ata aag gca tta gta gaa att tgt mca gaa ctg gaa atg gat gga Lys Ile Lys Ala Leu Val Glu Ile Cys Xaa Glu Leu Glu Met Asp Gly 432 aaa att tca aaa att ggg cct gaa aat ccg tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 480 gcc ata aag aaa aag aac agt act aaa tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Lys Asn Ser Thr Lys Trp Arg Lys Leu Val Asp Phe aga gaa ctt aac aaa aga act caa gac ttc tgg gaa gtt caa tta gga 576 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 185 ata cca cat ccc gca ggg tta aag aag aaa aaa tca gta aca gta ctg Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu 624

-7-

		195	5				200	0				205	5			
gat Asp	gtg Va] 210	r GTĀ	gat Asp	gca Ala	a tat a Tyr	ttt Phe 215	e Ser	a att	ccc Pro	tta Lev	tgt Cys 220	Glu	ı gad ı Asp	tto Phe	aga Arg	672
aag Lys 225	TAT	act Thr	gca Ala	ttt Phe	acc Thr 230	TIE	cct Pro	agt Ser	ata Ile	aac Asn 235	ı Asn	gag Glu	aca Thr	cca Pro	ggg Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	ASII	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
cag Gln	aat Asn	cca Pro 275	gaa Glu	atg Met	gtc Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gag Glu 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gat Asp	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	tat Tyr	ctg Leu	tgg Trp	aag Lys 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aca Thr	gaa Glu	PLO	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	TILL	gta Val 340	cag Gln	cct Pro	ata Ile	vaı	ctg Leu 345	cca Pro	gaa Glu	aaa Lys .	Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc a		gac a Asp 1 355	ata (cag Gln	aag Lys :	Leu	gtg Val 360	gga Gly	aaa Lys :	ttg Leu :	Asn '	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
<210: <211: <212: <213:	> 113 > DNA	A	mmur	odi	ficie	ency	Vin	1S (1	HIV)			•				
<220><221><222><223>	(0)	(297) teas	e												
<221><222><223>	(29	8)	.(11 of	16) HIV	Reve	rse	Tran	ıscri	.ptas	e						
<400> cct c Pro G	6 aa a	tc a	et e	tt t	aa c	aa c	~ ~		.		ca a hr I	ta a le L	ag a ys I	ta c	ily 199	48

-8-

1				5					10)				1	5	
99 G1	g ca y Gl	a ct n Le	a aa u Ly 2	s Gl	a gc u Al	t cta a Lei	a tta ı Lei	a gat 1 Ası 25	Thi	gga Gly	a gca y Ala	a gat a Asp	ga: Asj 3	p Th	a gta r Val	96
tt: ~~Lei	a gaa u Gli	a ga u As 3	p Me	g aat t Asi	t ttg n Le	g cca u Pro	gga Gly	Arc	tgg Trp	g aaa D Lys	a cca s Pro	a aaa o Lys 45	Met	g at	a ggg e Gly	144
G1,	a att 7 Ile 50	GT:	a ggi y Gly	t tti y Phe	t ato	c aaa e Lys 55	Val	agg Arg	g cag g Gln	tat Tyr	gat Asp 60	Glr	ata 111e	a cto	c ata u Ile	192
gaa Glu 65	ı Ile	tg: Cy:	t gga s Gly	a cat / His	aaa Lys 70	s Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	Lev	a gta 1 Val	gga Gly	cci	aca Thr 80	240
Pro	gto Val	aad Asi	ata 1 Ile	att E Ile 85	: Gly	a agg ⁄ Arg	aat Asn	ctg Leu	Leu 90	Thr	cag Gln	g att Ile	ggt	tgo Cys 95	act Thr	288
Leu	. Asn) Phe	Pro 100	Ile	: Ser	Pro	Ile	Glu 105	Thr	Val	Pro	Val	Lys 110	Leu	aag Lys	336
Pro	Gly	115	Asp	Gly	Pro	Lys	Val 120	Lys	Gln	Trp	Pro	Leu 125	Thr	Glu	gag Glu	384
ьуs	11e 130	Lys	Ala	Leu	Val	gaa Glu 135	Ile	Cys	Thr	Glu	Met 140	Glu	Lys	Glu	Gly	432
Lys 145	Ile	Ser	Lys	Ile	Gly 150	cct Pro	Glu	Asn	Pro	Tyr 155	Asn	Thr	Pro	Val	Phe 160	. 480
Ala	ITe	Lys	Lys	Lys 165	Asp	agt Ser	Thr	Lys	Trp 170	Arg	Lys	Leu	Val	Asp 175	Phe	528
Arg	GIU	Leu	Asn 180	Lys	Lys	act Thr	Gln	Asp 185	Phe	Trp	Glu	Val	Gln 190	Leu	Gly	576
TTE	Pro	H15 195	Pro	Ala	Gly		Lys 200	Lys	Lys	Lys	Ser	Val 205	Thr	Val	Leu	624
Asp	Val 210	GIY	Asp	Ala	Tyr	ttt Phe 215	Ser	Val	Pro	Leu	Asp 220	Lys	Asp	Phe	Arg	672
Lуs 225	ıyr	Inr	Ala	Phe	Thr 230	ata Ile	Pro	Ser	Ile .	Asn 235	Asn	Glu	Thr	Pro	Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr	aat Asn	gtg (Val :	ctt (Leu :	cca Pro	cag (Gln	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser	cca Pro	768

-9-

245 250 255 gca ata ttc caa agt agc atg ata aaa atc tta gag cct ttc aga aaa Ala Ile Phe Gln Ser Ser Met Ile Lys Ile Leu Glu Pro Phe Arg Lys 816 260 caa aat cca gac atg gtc atc tat caa tac atg gat gat ttg tat gta Gln Asn Pro Asp Met Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val 864 gga tot gao tta gaa ata gga cag cac aga aca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu 912 aga caa cat ctg ttg aag tgg gga ttt acc aca cca gac aag aaa cat Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His 960 cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 1008 aaa tgg aca gta cag cct ata aag ctg cca gaa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr 1056 340 gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 1104 att tac cca ggg Ile Tyr Pro Gly 1116 370 <210> 7 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV) <220> <221> CDS <222> (0)...(297) <223> HIV Protease <221> CDS <222> (298)...(1116) <223> Portion of HIV Reverse Transcriptase <400> 7 cct cag atc act ctt tgg caa cga ccc ctt gtc aca ata aar ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 48 ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 96 tta gag gaa atn aat tta cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Xaa Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 144

-10-

				35					4	0					4	5					
	99 G1	-	t g e G	ga d ly d	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	a ag l Ar	ac gG	ag (ln :	tat Tyr	gat Asp 60	Gl	gat n. I.	ca c le I	tt eu	gta Val	.	192
	6	5		•	,		70		116	GI	A TI	nr (75	Leu	Va.	L G1	y P	ro	Thr 80		240
						85		aga Arg	ASII	. Det	בי ייפ	90	nr	GIn	116	: Gl	y C	ys 95	Thr		288
				1	00		,	cct Pro	116	105	Th	ir v	aı.	Pro	Val	Ly 11	s Le O	≥u	Lys		336
			11	.5	JP C	-, -	10.		120	гуs	Hl	S G.	TÀ 1	Pro	Leu 125	Th:	r Gl	.u	Glu		384
	_	130	-		_		-	gaa Glu 135	116	Cys	m	r G.	Lu N	140	GLY	Lys	G1	u (Gly		432
	145			- - 2		1	50	ect o	31.0	ASI	Pro	15	7 A 55	sn '	Thr	Pro	Va	1 1	Phe L60		480
				1	16	55	JP C	gt a Ser I	. 111	ьys	170) Ar	gL	ys 1	Leu	Val	As _]	p E 5	Phe		528
	aga Arg			18	0 1		·			185	PILE	: II	pG.	ıu (/aI	GIn 190	Let	ı G	ly		576
	ata Ile		195			- 01	.y	2	00	uys	тўз	ьy	s Se	er V 2	05	Thr	Va]	. L	eu	•	624
	gat (Asp	210	_			1	21	L5	CT (aı .	PIO	ьег	22 22	sp G 20	lu A	Asp	Leu	G.	lu	•	572
	aag t Lys 1 225	-				230	5	.c Fi	.0 3	er.	тте	235	l As	n G	lu 1	hr	Pro	G: 24	ly 10	7	720
•	att a Ile A	_	-		245				.u F	20 0	50	Сту	1r	Ъг	ys G	TA	Ser 255	Pr	0	7	68
	gca a Ala I			260				C 111	20	65	те	ren	GI	u Pi	60 P 2	he . 70	Arg	Ьŷ	s	8	16
Ċ	aa a Sln A	at c	ro	gac Asp	ata Ile	gtt Val	ate Ile	c ta e Ty	t ca r G	aa t ln T	ac yr	atg Met	gat Asp	t ga	t t	tg :	tat Tyr	gt Va	a 1	8	64

-11-

	280	285	
gga tca gac tta gaa a Gly Ser Asp Leu Glu I 290	ta ggg čag cat aga le Gly Gln His Arg 295	aca aaa ata gag gaa ctg 912 Thr Lys Ile Glu Glu Leu 300	2
Arg Gln His Leu Leu G		aca cca gac aaa aaa cat 960 Thr Pro Asp Lys Lys His 315 320)
cag aaa gaa cct cca t Gln Lys Glu Pro Pro P 325	tc ctt tgg atg ggt he Leu Trp Met Gly 330	tat gaa ctc cat cct gat 1008 Tyr Glu Leu His Pro Asp 335	3
aaa tgg aca gta cag c Lys Trp Thr Val Gln P 340	ct ata gtg ctg cca ro Ile Val Leu Pro . 345	aca aaa gac agc tgg act 1056 Thr Lys Asp Ser Trp Thr 350	i
gtc aat gac ata cag a Val Asn Asp Ile Gln L 355	ag tta gtg gga aaa ys Leu Val Gly Lys 360	ttg aac tgg gca agt cag 1104 Leu Asn Trp Ala Ser Gln 365	:
att tat gca ggg Ile Tyr Ala Gly 370		1116	ï
<210> 8 <211> 1116 <212> DNA <213> Human Immunodif: <220> <221> CDS <222> (0)(297) <223> HIV Protease	iciency Virus (HIV)		
<221> CDS <222> (298)(1116) <223> Portion of HIV R	Reverse Transcriptas	use	
<400> 8 cct cag atc act ctt tg Pro Gln Ile Thr Leu Tr 1 5	gg caa cga ccc cty o p Gln Arg Pro Xaa v 10	gtc aca gta aag ata ggg 48 Val Thr Val Lys Ile Gly 15	
cct cag atc act ctt tg Pro Gln Ile Thr Leu Tr 1 5	p Gln Arg Pro Xaa v 10 t yta tta qat aca o	Val Thr Val Lys Ile Gly	
cct cag atc act ctt tg Pro Gln Ile Thr Leu Tr 1 5 ggg caa ata aag gaa gc Gly Gln Ile Lys Glu Al 20 tta gaa gaa atg aat tt	tyta tta gat aca ca Xaa Leu Asp Thr Ca 25	Val Thr Val Lys Ile Gly 15 gga gca gat gat aca gta Gly Ala Asp Asp Thr Val	
cct cag atc act ctt tg Pro Gln Ile Thr Leu Tr 1 5 ggg caa ata aag gaa gc Gly Gln Ile Lys Glu Al 20 tta gaa gaa atg aat tt Leu Glu Glu Met Asn Le 35	tyta tta gat aca ga Xaa Leu Asp Thr Cass ga aga tgg aca gga aga tgg aca Pro Gly Arg Trp I	Val Thr Val Lys Ile Gly 15 gga gca gat gat aca gta Gly Ala Asp Asp Thr Val 30 aaa cca aaa ata ata ggg Lys Pro Lys Ile Ile Gly	

-12-

65	5				70)				75	5				80	
cct Pro	gto Val	aat L Asr	ata 11e	a att	≥ Gly	a aga / Arg	a aat g Asr	cto Lev	g atg Met 90	Thr	cag Glr	g att	ggt Gly	tgo Cys	act Thr	288
tta Leu	aat Asn	ttt Phe	Pro) Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	Asp	ggc Gly	Pro	aga Arg	gtt Val 120	. Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	Glu	aag Lys	gaa Glu	Gly	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	Gly	agt Ser	aac Asn	aga Arg	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aaa Lys	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	Gly 999	cta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gly aaa	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
Ser	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	cta Leu	912
aga Arg	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg	tgg Trp	gga Gly	tta Leu	acc Thr	aca Thr	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His	960

-13-

	310	315	320
cag aaa gaa cct cca Gln Lys Glu Pro Pro 325	ttc ctt tgg atg ggt Phe Leu Trp Met Gly 330	tat gaa ctc cat cct Tyr Glu Leu His Pro 339	Asp
aaa tgg aca gta cag _Lys Trp Thr Val Gln 340	cct ata gtg ctg cca Pro Ile Val Leu Pro 345	gar aaa gac agc tgc Glu Lys Asp Ser Trp 350	act 1056 Thr
gtc aat gac ata cag Val Asn Asp Ile Gln 355	aag tta gtg gga aaa Lys Leu Val Gly Lys 360	ttg aat tgg gca agt Leu Asn Trp Ala Ser 365	caa 1104 Gln
att tac cca ggg Ile Tyr Pro Gly 370			1116
<210> 9 <211> 1116 <212> DNA <213> Human Immunodi	ficiency Virus (HIV)		
<220> <221> CDS <222> (0)(297) <223> HIV Protease			
<221> CDS <222> (298)(1116)			
<223> Portion of HIV	Reverse Transcripta	ase	
<223> Portion of HIV <400> 9 cct cag atc act ctt Pro Gln Ile Thr Leu 1 5	tgg caa cga ccc ctv	gtc aaa gta aag ata	ggg 48 Gly
<400> 9 cct cag atc act ctt Pro Gln Ile Thr Leu	tgg caa cga ccc cty Trp Gln Arg Pro Xaa 10 gct cta tta gat aca	gtc aaa gta aag ata Val Lys Val Lys Ile 15	Gly
<pre><400> 9 cct cag atc act ctt Pro Gln Ile Thr Leu 1</pre>	tgg caa cga ccc cty Trp Gln Arg Pro Xaa 10 gct cta tta gat aca Ala Leu Leu Asp Thr 25	gtc aaa gta aag ata Val Lys Val Lys Ile 15 gga gca gat gat aca Gly Ala Asp Asp Thr 30 aaa cca aaa atg ata	gta 96 Val
<pre><400> 9 cct cag atc act ctt Pro Gln Ile Thr Leu 1</pre>	tgg caa cga ccc cty Trp Gln Arg Pro Xaa 10 gct cta tta gat aca Ala Leu Leu Asp Thr 25 ctg cca gga aga tgg Leu Pro Gly Arg Trp 40	gtc aaa gta aag ata Val Lys Val Lys Ile 15 gga gca gat gat aca Gly Ala Asp Asp Thr 30 aaa cca aaa atg ata Lys Pro Lys Met Ile 45	gta 96 Val ggg 144 Gly
<pre><400> 9 cct cag atc act ctt Pro Gln Ile Thr Leu 1</pre>	tgg caa cga ccc cty Trp Gln Arg Pro Xaa 10 gct cta tta gat aca Ala Leu Leu Asp Thr 25 ctg cca gga aga tgg Leu Pro Gly Arg Trp 40 atc aaa gta aga cag Ile Lys Val Arg Gln 55	gtc aaa gta aag ata Val Lys Val Lys Ile 15 gga gca gat gat aca Gly Ala Asp Asp Thr 30 aaa cca aaa atg ata Lys Pro Lys Met Ile 45 tat gat cag ata ccc Tyr Asp Gln Ile Pro 60	gta 96 Val ggg 144 Gly ata 192 Ile
<pre><400> 9 cct cag atc act ctt Pro Gln Ile Thr Leu 1</pre>	tgg caa cga ccc cty Trp Gln Arg Pro Xaa 10 gct cta tta gat aca Ala Leu Leu Asp Thr 25 ctg cca gga aga tgg Leu Pro Gly Arg Trp 40 atc aaa gta aga cag Ile Lys Val Arg Gln 55 aaa gct ata ggt aca Lys Ala Ile Gly Thr 70	gtc aaa gta aag ata Val Lys Val Lys Ile 15 gga gca gat gat aca Gly Ala Asp Asp Thr 30 aaa cca aaa atg ata Lys Pro Lys Met Ile 45 tat gat cag ata ccc Tyr Asp Gln Ile Pro 60 gta tta gta gga cct Val Leu Val Gly Pro 75	gta 96 Val ggg 144 Gly ata 192 Ile . aca 240 Thr 80

-14-

			100)				105	5				110)		
Pro	a gga o Gly	a ato / Met 115	: Asp	ggo Gly	cca Pro	a aaa D Lys	gtt Val	Lys	caa Glr	tgg Trp	p cca	tto Lev 125	Thi	a gaa Glu	a gaa 1 Glu	384
aaa Lys	a ata s Ile 130	: Lys	a gca s Ala	i tta i Lev	ata Ile	gaa Glu 135	Ile	tgt Cys	aca Thr	gag Glu	ato Met 140	: Glu	g aa⊆ Lys	gaa Glu	a ggg a Gly	432
aaa Lys 145	: Ile	tca Ser	aaa Lys	att Ile	ggg Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc	ata Ile	aag Lys	aaa Lys	aaa Lys 165	Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	. Lys	aaa Lys	act Thr	caa Gln	gay Asp 185	Phe	tgg Trp	gaa Glu	gtt Val	car Gln 190	Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	Pro	gca Ala	gly aaa	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	. 672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aag Lys 265	atc Ile	tta Leu	gar Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gaa Glu	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tcw Xaa 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	caa Gln	cat His	aga Arg	ata Ile	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	cag Gln	cat His	ctg Leu	tta Leu	agg Arg 310	tgg Trp	Gly 999	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggk Xaa 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lvs	tgg Tro	aca Thr	gta Val	cag Gln	CCT	ata Ile	gtg Val	ctg	cca	gaa	aaa Lve	gay	agc	tgg	act	1056

-15-

				34	U				34	5				35	O		
	gt Va	c aa l As	t ga n As 35	p Il	a ca e Gl	g aa n Ly	g tta s Le	a gte u Val 360	l Gl	a aa y Ly	a tt	g aa u As:	t tgg n Trj 36!	p Al	a ag a Se:	t cag r Gln	1104
_	ate Ile	c tag e Ty: 37	c cc r Pro	a gg o Gl	g Y												1116
	<2:	10> 1 11> 1 12> I 13> I	1116 ONA	ı Imi	nuno	dific	cienc	y Vi	.rus	(HIV	7)						
	<22	21> (22> (DS (0)	(29 Prote	97) ease		•								-		
	<22		(298)		(1116 of H		vers	e Tr	ansc	ript	ase						
	cct	0> 1 cag Gln	ato	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc	cto Leu 10	Val	aca Thr	gta Val	aag Lys	ata Ile 15	ggg	48
	Gly ggg	caa Gln	ata Ile	aag Lys 20	Glu	gct Ala	yta Xaa	tta Leu	gat Asp 25	Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gta Val	96
	tta Leu	gaa Glu	gaa Glu 35	Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atw Xaa	ata Ile	gjå aaa	144
	gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	gta Val	ccc Pro	ata Ile	192
	gaa Glu 65	atc Ile	tgt Cys	gga Gly	caa Gln	aaa Lys 70	gct Ala	ata Ile	agt Ser	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
	cct Pro	gtc Val	aat Asn	ata Ile	att Ile 85	Gly	aga Arg	aat Asn	ctg Leu	atg Met 90	Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	288
	tta Leu	aat Asn	ttt Phe	cct Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	taa ☆	aag Lys	336
	cca Pro	gga Gly	atg Met	gat Asp 115	ggc	cca Pro	aga Arg	gtt Val	aaa Lys 120	caa Gln	tgg Trp	cca Pro	ttg Leu	aca Thr 125	gaa Glu	gaa Glu	384
	aaa Lys	ata Ile	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu	atc Ile	tgt Cys	aca Thr	gaa Glu	atg Met	gaa Glu	aag Lvs	gaa Glu	ggg Glv	432

-16-

		130					135					140				
aaa Lys	att Ile 145	tca Ser	aaa Lys	att Ile	gly	cct Pro 150	gaa Glu	aat Asn	cca Pro	tac Tyr	aat Asn 155	act Thr	cca Pro	gta Val	ttt Phe	480
gcc Ala 160								aga Arg								528
								gac Asp								576
ata Ile	cca Pro	cat His	ccc Pro 195	gca Ala	Gly	cta Leu	aaa Lys	aag Lys 200	aaa Lys	aaa Lys	tca Ser	gta Val	aca Thr 205	gta Val	ctg Leu	624
								gtt Val								672
aag Lys	tat Tyr 225	act Thr	gca Ala	ttt Phe	acc Thr	ata Ile 230	cct Pro	agt Ser	aca Thr	aac Asn	aat Asn 235	gag Glu	aca Thr	cca Pro	Gly 999	720
								ccm Xaa								768
								aaa Lys								816
								caa Gln 280								864
								cat His								912
								tta Leu								960
								atg Met								1008
aaa Lys	tgg Trp	aca Thr	gta Val	cag Gln 340	cct Pro	ata Ile	gtg Val	ctg Leu	cca Pro 345	gag Glu	aaa Lys	gac Asp	agc Ser	tgg Trp 350	act Thr	1056
gtc Val																1104
att Ile																1116

-17-

370

<2 <2	12>	1116 DNA	n Im	muno	difi:	cien	cy V	irus	(HI	V)						
<2 <2		(0).	(2: Prot													
<2	21> 22> 23>	(298) ion ((1110 of H:	5) IV Re	evers	se Ti	ranso	ript	ase						
cci	00> : t cag c Gl	gate	c act e Thi	ctt Leu 5	tgg Trp	g caa Glr	a cga	a ccc	aty Xaa	ı Val	aca Thr	ata : Ile	a aag E Lys	g ata s Ile 15	ggg Gly	4.8
G1 ³	g caa / Gli	a cta n Lei	a aaa 1 Lys 20	: Xaa	gct Ala	cta Leu	tta Lev	a gat Asp 25	Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gta Val	96
tta Lev	gaa Glu	gaa Glu 35	atg Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	' Arg	tgg	aaa Lys	cca Pro	aaa Lys 45	Met	ata Ile	gtg Val	144
gga Gly	att Ile 50	: GLY	ggt Gly	ttt Phe	gtc Val	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	Gln	gta Val	ccc	ata Ile	192
gag Glu 65	TTE	tgt Cys	G1 ggg	cat His	aaa Lys 70	att Ile	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	ata Ile	gga Gly	cct Pro	acc Thr 80	. 240
cct Pro	gcc Ala	aac Asn	gta Val	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	atg Met 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	yct Xaa	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	gca Ala	gaa Glu	ctg Leu 140	gag Glu	aag Lys	gaa Glu	ggg Gly	432
aaa Lys 145	att Ile	tca Ser	aga Arg	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	ata Ile	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aag Lys	aaa Lys	aac Asn	agt Ser	act Thr	agg Arg	tgg Trp	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp	ttc Phe	528

-18-

				165					170					175		
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	att Ile	caa Gln 190	tta Leu	gga Gly	576
ata ~~Ile	cca Pro	cat His 195	cct Pro	gca Ala	Gly 999	tta Leu	aaa Lys 200	aag Lys	aac Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	Gly 999	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tac Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	gat Asp	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gat Asp	ccc Pro	ttt Phe 270	aga Arg	aag Lys	816
aaa Lys	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	yta Xaa	gaa Glu	ata Ile	gag Glu 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	gaa Glu	tat Tyr	ctg Leu	tta Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	ttc Phe	aca Thr 315	cca Pro	gag Glu	caa Gln	aaa Lys	cat His 320	960
Gln	Lys	Ğlu	Pro	Pro 325	Phe	Leu	Trp	atg Met	Gly 330	Tyr	Glu	Leu	His	Pro 335	Asp	1008
aaa Lys	tgg Trp	aca Thr	gtg Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aac Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	
	tac Tyr 370															1116

<210> 12 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)

-19-

```
<220>
  <221> CDS
  <222> (0)...(297)
<223> HIV Protease
  <221> CDS
  <222> (298)...(1116)
-~<223> Portion of HIV Reverse Transcriptase
 cct caa atc act ctt tgg car cga ccc tta gtc aca ata aag ata ggg
Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
                                                                                            48
                                                                                            96
  ggg caa cta aag gaa gcc cta tta gat aca gga gca gat gat aca gta
 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
  cta gaa gaa atg aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg
                                                                                           144
 Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly
 gga att gga ggt ttt atc aaa gta agg cag tat gat car ata ccc ata
                                                                                           192
 Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
 gag atc tgc ggg tat aaa gct gtg ggt aca gta tta gta gga cct aca Glu Ile Cys Gly Tyr Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr
                                                                                           240
 cct gtc aac ata att gga aga aat ctg ttg act caa att ggt tgc act
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                                                                                           288
 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
                                                                                           336
                 100
 cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
                                                                                           384
                                      120
                                                                                           432
 aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa gga
 Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
 aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt
                                                                                           480
 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
 145
 gcc ata aag aaa aaa gac ggt act aaa tgg aga aaa tta gta gat ttc
                                                                                           528
 Ala Ile Lys Lys Lys Asp Gly Thr Lys Trp Arg Lys Leu Val Asp Phe
                                                                                           576
 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga
 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
                                           185
 ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta cta
Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu
                                                                                           624
```

-20-

		195					200					205				
		Gly				ttt Phe 215						Gln			aga Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttc Phe	act Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	atg Met	gtt Val	atc Ile	tat Tyr 280	caa Gln	tat Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
ggc Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	aya Xaa	999 Gly 295	cag Gln	cat His	aga Arg	rca Xaa	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttt Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	atg Met	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	cta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile																1116
<210 <211 <212 <213	> 11 > DN	A	Immu	nodi	fici	ency	Vir	us (HIV)							
<220: <221: <222: <223:	> CD; > (0)	·														
<221: <222: <223:	> (2	98).			Rev	erse	Tra	nscr	ipta	se						

-21-

		0> 1:															4.0
	Pro 1	cag Gln	atc Ile	act Thr	ctt Leu 5	tgg Trp	caa Gln	Arg	Pro	Xaa 10	gtc Val	aac Asn	Ile	aag Lys	Val 15	ggg Gly	48
-4	Gly	caa Gln	cta Leu	arg Xaa 20	gaa Glu	gct Ala	cta Leu	ata Ile	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
	tta Leu	gaa Glu	gac Asp 35	ata Ile	gat Asp	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aga Arg	cca Pro	aga Arg 45	atg Met	ata Ile	Gly 999	144
	gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	gtc Val	aaa Lys 55	gta Val	aag Lys	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	ccc Pro	ata Ile	192
	gaa Glu 65	ata Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gtt Val	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	acg Thr 80	240
	cct Pro	gcc Ala	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	Gly aaa	tgc Cys 95	act Thr	288
	tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aaa Lys	336
	cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
	aag Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
	aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
	gcc Ala	ata Ile	aag Lys	aag Lys	aaa Lys 165	aac Asn	agt Ser	act Thr	aga Arg	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttt Phe	528
	aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttt Phe	tgt Cys	gaa Glu	gtg Val	caa Gln 190	tta Leu	gga Gly	576
	ata Ile	ccg Pro	cat His 195	ccc Pro	gca Ala	gly aaa	tta Leu	ara Xaa 200	aag Lys	aaa Lys	aga Arg	tca Șer	gta Val 205	aca Thr	gta Val	ctg Leu	624
	gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
	aag Lys 225	tat Tyr	act Thr	gcc Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720

-22-

	att Ile	aga Arg	tat Tyr	cag Gln	tat Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
									aaa Lys 265								816
									caa Gln								864
	gly ggg	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	caa Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gag Glu	ttg Leu	912
	aga Arg 305	gaa Glu	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	tty Phe	ttc Phe	aca Thr 315	cca Pro	gac Asp	gaa Glu	aaa Lys	cat His 320	960
									atg Met								1008
									ttg Leu 345								1056
									gga Gly								1104
		tac Tyr 370														•	1116
	<211 <212	> 14 > 11 > DN > Hu	16 A	Immu	nodi	fici	ency	· Vir	rus (HIV)							
	<222	> CD > (0)	(297 otea	-						·						
	<222		98).	(1 n of		Rev	erse	Tra	nscr	ipta	se						
	cct	> 14 caa Gln	atc Ile	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	ctc Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	gly ggg	48
•	31y (caa Gln	gta Val	agg Arg 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96

	a gaa u Glu		ı Met					Lys					Met			144
99 G1	a att y Ile 50	e Gly	ggg Gly	ttt Phe	ato lle	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	Gln	ata Ile	ccc Pro	ata Ile	192
	a ato u Ile 5					Āla					Leu					240
	t gto o Val				Gly											288
tt: Le:	a aat u Asn	ttt Phe	Pro	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gtg Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
	a gga o Gly		Asp													384
aaa Lys	a ata s Ile 130	Lys	gca Ala	ttg Leu	ata Ile	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aag Lys 145	g att G Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gco	ata Ile	aag Lys	aaa Lys	aaa Lys 165	aac Asn	agt Ser	act Thr	aga Arg	tgg Trp 170	agg Arg	aaa Lys	cta Leu	gta Val	gac Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	gga Gly	tta Leu	aaa Lys 200	aag Lys	aga Arg	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aar Lys	gac Asp	ttc Phe	agg Arg	672
	tac Tyr															720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctg Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
tca Ser	ata Ile	ttc Phe	caa Gln 260	agt Ser	agy Xaa	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aag Lys	816

WO 01/35316 PCT/US00/30863

-24-

ca Gl	a aat n Asn	cca Pro 275	gac Asp	ata Ile	gat Asp	atc Ile	tgt Cys 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
	a tct y Ser 290															912
ag Ar 30	a gag g Glu 5	cat His	ctg Leu	cta Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	raa Xaa	aaa Lys	cat His 320	960
	r aaa n Lys															1008
	a tgg s Trp															1056
gt: Va	aat l Asn	gac Asp 355	ata Ile	caa Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
	tat Tyr 370												•			1116
<23 <23	10> 15 11> 11 12> DN	l16 IA	Immu	ınodi	fici	.ency	/ Vii	cus ((HIV)							
<22	20> 21> CI 22> (0 23> HI))										•				
<22	21> CI 22> (2 23> Po	98).				erse	e Tra	ınscr	ripta	ıse						
cct	00> 15 caa Gln	atc	act Thr	ctt Leu 5	tgg Trp	car Gln	cga Arg	ccc Pro	ctc Leu 10	gtt Val	gca Ala	ata Ile	aag Lys	ata Ile 15	Gly 999	48
	caa Gln															96
tta Leu	kaa Xaa	gaa Glu 35	atg Met	gat Asp	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly 999	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	gta Val	tcc Ser	wta Xaa	192

	Ile				aaa Lys 70	Ala					Leu				aca Thr 80	240
					gga Gly					Thr					Thr	288
				Ile	agt Ser									Leu		336
			Āsp		cca Pro			Lys					Thr			384
		Lys			gta Val											432
	Ile				999 Gly 150											480
					gac Asp											528
					aaa Lys											576
					gly ggg											624
					tat Tyr											672
					acc Thr 230											720
					aat Asn											768
					agc Ser											816
caa Gln	aat Asn	cca Pro 275	gaa Glu	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gat Asp	tta Leu	gaa Glu	ata Ile	gaa Glu 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912

-26-

305 310 315 320	960
cag aaa gaa cct cca ttc ctt tgg atg ggk tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Xaa Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag ttr gtg gga aaa ttr aat tgg gca agt cag Val Asn Asp Ile Gln Lys Xaa Val Gly Lys Xaa Asn Trp Ala Ser Gln 355 360 365	1104
att tac tca ggg Ile Tyr Ser Gly 370	1116
<210> 16 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS	
<222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><222> (298)(1116) <223> Portion of HIV Reverse Transcriptase <400> 16 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
<223> Portion of HIV Reverse Transcriptase <400> 16 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly	48 96
<pre><223> Portion of HIV Reverse Transcriptase <400> 16 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1</pre>	
<pre><223> Portion of HIV Reverse Transcriptase <400> 16 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1</pre>	96
<pre><223> Portion of HIV Reverse Transcriptase <400> 16 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1</pre>	96 144

tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
 cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	rta Xaa	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aag Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	4.80
gcc Ala	ata Ile	aag Lys	aaa Lys	aar Lys 165	gat Asp	ggt Gly	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	att Ile	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	gly ggg	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aag Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	act Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	tgt Cys	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aag Lys	816
caa Gln	aat Asn	Pro	Asp	Ile	gtt Val	Ile	Tyr	${\tt Gln}$	Tyr	Met	Asp	Asp	Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	rca Xaa	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
agg Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cca Pro 335	gat Asp	1008

-28-

aaa tgg aca gta cag cct ata gtg ctg cca caa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Gln Lys Asp Ser Trp Thr 340 345 350	1056										
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104										
att tat cca ggg Ile Tyr Pro Gly 370	1116										
<210> 17 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)											
<220> <221> CDS <222> (0)(297) <223> HIV Protease											
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase											
<400> 17 cct caa atc act ctt tgg caa cga ccc aty gtc aca ata aag gta ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Val Gly 1 5 10 15	48										
ggg caa cta aag gaa gcc cta ata gat aca gga gca gat gat aca gtg Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96										
tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa ttg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly 35 40 45	144										
gga att gga ggt ttt atc aaa gta aga cag tat gat cag rta ccc ata Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Xaa Pro Ile 50 55 60	192										
gaa atc tgt gga cat aaa gct gta ggt tca gtg tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Val Gly Ser Val Leu Val Gly Pro Thr 65 70 75 80											
cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 85 90 95	288										
cta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336										
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca aaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Lys Glu 115 120 125	384										

aaa Lys	ata Ile 130	gaa Glu	gca Ala	tta Leu	gta Val	gaa Glu 135	atc Ile	tgt Cys	gca Ala	gaa Glu	ctg Leu 140	gaa Glu	gag Glu	gca Ala	gly ggg	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aar Lys	aag Lys 165	aac Asn	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aac Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	Gly 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	att Ile	ccc Pro	tta Leu	gat Asp 220	aag Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	aca Thr 230	ata Ile	cct Pro	agy Xaa	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cma Xaa	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	cag Gln 260	tgt Cys	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gat Asp	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	car Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	Gln	Xaa	Leu	\mathtt{Trp}	Lys	\mathtt{Trp}	Gly	ttt Phe	Tyr	Thr	cca Pro	gag Glu	aat Asn	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	cwt Xaa	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aaa Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104

-30-

att tat gcn ggg Ile Tyr Ala Gly 370	1116											
<210> 18 _<211> 1117 <212> DNA <213> Human Immunodificiency Virus (HIV)												
<220> <221> CDS <222> (0)(297) <223> HIV Protease												
<221> CDS <222> (298)(1117) <223> Portion of HIV Reverse Transcriptase												
<pre><400> 18 cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48											
ggg car cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96											
gta gaa gaa atg aat tta tca gga agg tgg aaa cca aaa atg ata ggg Val Glu Met Asn Leu Ser Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144											
gga att gga ggt ttt atc aaa gta aga saa tat gaa cag ata cct gta Gly Ile Gly Gly Phe Ile Lys Val Arg Xaa Tyr Glu Gln Ile Pro Val 50 55 60	192											
gaa att tgt gga cat aaa gct gta ggt aca gta tta gtg gga cct aca Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240											
cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 85 90 95	288											
tta aat ttt ccc att agt ccc att gaa act gta cca gta aaa ttg aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336											
cca gga atg gat ggc ccg aga gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384											
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135 140	432											
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 145 150 155 160	480											

-31-

go Al	c at a Il	a aa e Ly	ag a ys L	ys 1	aaa Lys 165	gac Asp	agt Ser	aat Asn	aaa Lys	tgg Trp 170	agg Arg	aaa Lys	tta Leu	gtg Val	gat Asp 175	ttc Phe	528
ac Ar	ga ga g Gl	a ct u Le	eu A	at a sn 1 80	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	Gly aaa	576
at Il	a cc e Pr	a ca o Hi	is X	cy i	tca Ser	Gly 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
ga As	t gt p Va 21	1 G1	gt ga Ly Aa	at q sp 1	gca Ala	tac Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg	672
aa Ly 22	s Ty	t ac r Th	et ge	ca t la 1	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
at Il	t ag e Xa	r ta a Ty	at ca /r G	ln 7	tac Tyr 245	aat Asn	gtg Val	ctg Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gc Al	a at a Il	a tt e Ph	ie G	aa a ln S 60	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	gaa Glu	816
ca Gl	a aa n As:	t ac n Th 27	ır As	ac a sp 1	ata Ile	gtt Val	atc Ile	tgt Cys 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gg G1	a tc y Se 29	r As	c ti p Le	ta g eu (gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	gtr Xaa	gag Glu	gaa Glu	ctg Leu	912
ag Ar 30	g Gl	a ca n Hi	t ct .s Le	tg t eu I	tg Leu	agg Arg 310	tgg Trp	gga Gly	yta Xaa	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
				ro E										cat His			1008
			ır Xa											agc Ser 350			1056
gt Va	c aat l Ası	ga 1 As 35	p Il	ta d le G	caa Sln	aaa Lys	gtt Val	agt Ser 360	gly 999	aaa Lys	att Ile	aaa Lys	ttg Leu 365	ggc Gly	aag Lys	tca Ser	1104
	t tta p Lei 370	ı Pr			3												1117

<210> 19 <211> 1116 -32-

```
<212> DNA
<213> Human Immunodificiency Virus (HIV)
<220>
<221> CDS
<222> (0)...(297)
<223> HIV Protease
<221> CDS
<222> (298) ... (1116)
<223> Portion of HIV Reverse Transcriptase
cct cag atc act ctt tgg caa cga ccc cty gtc aca gta aag ata ggg
Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Val Lys Ile Gly
                                                                                       48
ggg caa cta acg gaa gct yta ttg gat aca gga gca gat aat aca gta
Gly Gln Leu Thr Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asn Thr Val
                                                                                       96
                20
tta gaa gaa atg agt ttr cca gga aga tgg aaa cca aaa atg ata ggg
Leu Glu Glu Met Ser Xaa Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
                                                                                      144
                                                                                      192
gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
gaa atc tgt gga cat aaa gta gta ggt aca gta tta ata gga cct aca
Glu Ile Cys Gly His Lys Val Val Gly Thr Val Leu Ile Gly Pro Thr
                                                                                      240
                                                                                      288
ect gtc aac ata att gga aga gat ctg ttg act cag att ggt tgc act
Pro Val Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Ile Gly Cys Thr
tta aat ttt ccc att agt cct att gaa act gta cca gtg aaa tta aag
                                                                                      336
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
                                        105
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
                                                                                      384
aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gaa ggg
                                                                                      432
Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly
                                                                                      480
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
                         150
                                                                                      528
gcc ata aag aaa aar gac agt act aaa tgg aga aaa ttr gta gat ttc
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Xaa Val Asp Phe
aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt caa tta gga
                                                                                      576
Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly
```

-33-

	ata Ile	cca Pro	cat His 195	ccc Pro	tca Ser	Gly 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	cta Leu	624
14	gac Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	gtt Val	ccc Pro	cta Leu	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg	672
	aag Lys 225	tat Tyr	act Thr	gca Ala	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aac Asn 235	aat Asn	gag Glu	act Thr	cca Pro	999 Gly 240	720
	att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctg Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
	gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
	cac His	aat Asn	cca Pro 275	aac Asn	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	tta Leu	tat Tyr	gta Val	864
	gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	gta Val	gag Glu	gaa Glu	ctg Leu	912
	aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	ggg ggg	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
	cag Gln	aaa Lys	gaa Glu	ccc Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
	aaa Lys	tgg Trp	aca Thr	gtg Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	Asp	agc Ser 350	tgg Trp	act Thr	1056
	gtc Val	aat Asn	gac Asp 355	ata Ile	caa Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
	Ile	tac Tyr 370															1116
	<211 <212)> 20 .> 11 !> DN !> Hu	.17 IA	Immu	modi	.fici	ency	, Vir	rus ((HIV)							
	<222	> CI > (0))	. (297 cotea													

-34-

<221> CDS <222> (298)...(1117) <223> Portion of HIV Reverse Transcriptase <400> 20 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata gga 48 Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta 96 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val tta gaa gac ata aat ttg cca ggg aaa tgg aaa cca aaa atg ata ggg Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly 144 gga att gga ggt ttt atc aaa gta aga cag tat gat caa ata cca gta Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Val 192 240 gaa att tgt gga cat aaa gct gta ggt aca gta tta ata gga cct aca Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Ile Gly Pro Thr cct gtc aac gta att gga aga aat ctg atg act cag att ggc tgc act 288 Pro Val Asn Val Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr 90 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag 336 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys cca gga atg gat ggt cca aaa gtt aaa caa tgg cca tta aca gaa gaa 384 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 120 aaa ata aaa gca tta gta gaa att tgc aca gaa ttg gaa aag gaa ggg 432 Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly aaa att tca aaa att gga cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 480 gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe 528 165 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga 576 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 185 ata cca cat cct gca ggg tta aaa aag aaa aaa tca gta aca gta ctg 624 Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu 195 gat gtg ggt gat gca tat ttt tca ata ccc tta gat gaa gaa ttc agg 672 Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Glu Glu Phe Arg

-35-

aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	cca Pro	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttt Phe	caa Gln 260	tgt Cys	agt Ser	atg Met	Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
gaa Glu	aat Asn	cca Pro 275	gat Asp	ata Ile	gtt Val	atc Ile	tac Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gac Asp 285	tta Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gat Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	tat Tyr	ctg Leu	tgg Trp	aag Lys 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	caa Gln	gaa Glu	cct Pro	cca Pro 325	ttc Phe	cgt Arg	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg T r p	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	ttt Phe	agt Ser 360	Gly ggg	aaa Lys	att Ile	gaa Glu	ttg Leu 365	ggc Gly	aag Lys	tca Ser	1104
	tta Leu 370			g												1117
<21:	0> 21 L> 11 2> DN 3> Hu	l16 JA	Imm	ınodi	ifici	iency	y Viz	rus ((HIV)							
<222	0> L> CI 2> (0 3> HI))														
<222	L> CI 2> (2 3> Po	298).				verse	e Tra	nsci	ripta	ase						
cct	0> 21 cag Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	mcc Xaa	gtt Val 10	gtc Val	wca Xaa	ata Ile	aag Lys	ata Ile 15	Gly 999	48

99	g caa	cta	aaa	gaa	gct	cta	tta	gay	aca	āāa	gça	gat	gat	aca	gta	96
Gl;	y Gln	Leu	Lys 20	Glu	Ala	Leu	Leu	Asp 25	Thr	GIY	ALA	Asp	30	Thr	vaı	
tt. Le	a gaa ı Glu	gac Asp 35	atg Met	cat His	ttg Leu	cca Pro	ggt Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	gtg Val	144
gg: Gl:	a att 7 Ile 50	ggg Gly	ggt Gly	ttt Phe	gtc Val	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	cct Pro	gta Val	192
ga: Gl: 6!	a atc 1 Ile 5	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
Pro	a gcc o Ala	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	288
tta Le:	a aat u Asn	ttc Phe	ccc Pro 100	atc Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
Pro	a gga o Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	att Ile 120	aga Arg	caa Gln	tgg Trp	cca Pro	tta Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	a ata s Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	atc Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	Gly ggg	432
aaa Lys 145	att s Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gc Ala	ata lle	aag Lys	aaa Lys	aaa Lys 165	aat Asn	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	Gly aaa	Leu	aaa Lys 200	Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	\mathtt{Thr}	gta Val	cta Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	atg Met	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	gga Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	atg Met 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768

-37-

	gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agt Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
-1:	Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtc Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	tta Leu	tat Tyr	gta Val	864
	gga Gly	tcg Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	ggg Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ttg Leu	912
	Arg 305	Gln	His	Leu	Leu	310	пр	GIY	ttt Phe	1111	315			-3 -	•	320	960
	cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
	aaa Lys	tgg Trp	aca Thr	gtg Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
	gtt Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	caa Gln	1104
		tat Tyr 370															1116
	<21	0 > 2 1 > 1 2 > D 3 > H	116 NA	Imm	unod	ific	ienc	y Vi	rus	(HIV	·)						
	<22	0> 1> C 2> (3> H	0)														
	-22	1> C 2> (3> P	298)	(on o	1116 f HI) V Re	vers	e Tr	ansc	ript	ase					·	
		0> 2 cag		act Thr	ctt Lev 5	tgg Trp	caa Glr	cga Arg	ccc Pro	cto Leu 10	r val	acs Thr	ata : Ile	aag Lys	gta Val	gga Gly	48
	Gly 999	caa Gln	. cta Leu	aag Lys 20	GIU	gct Ala	cta Lev	tta Lev	gat Asp 25	1111	gga Gly	a gca / Ala	a gat a Asp	gat Asp 30		gta Val	96
	tta Lev	gaa Glu	gac Asp 35	Ile	gat Asp	tto Lev	g cca	a gga o Gly 40	Add	tgg	g aaa b Lys	a cca s Pro	a aaa Lys 45		g ata : Ile	e Gly	144

-38-

gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	ccc Pro	ata Ile	192
gaa Glu 65	ata Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cgg Arg	att Ile	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttt Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gat Asp 185	ttc Phe	tgg Trp	gaa Glu	gtg Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	ccg Pro	cat His 195	ccc Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aar Lys	gay Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gcc Ala	Phe	Thr	Ile	Pro	Ser	Ile	Asn	aat Asn	Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tat Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	cta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	atg Met	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864

-39-

ggg tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu 290 295 300	912
aga gaa cat ctg ttg agg tgg gga ttt acc acc cca gac aaa aaa cat Arg Glu His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His 305 310 315 320	960
cag aaa gag cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg acc gtr cag cct ata gag ctg cca gaa aaa gac agc tgg act Lys Trp Thr Xaa Gln Pro Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tac cca ggg Ile Tyr Pro Gly 370	1116
<210> 23 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 23 cct cag atc act ctt tgg caa cga ccc ata gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta aag gaa gct cta ata gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gac ata aat ttg cca gga aga tgg aaa cca aaa tta ata ggg Leu Glu Asp Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly 35 &0 &5	144
gga att gga ggt ttt gtc aga gtg aaa cag tat gat cag ata ccc ata Gly Ile Gly Gly Phe Val Arg Val Lys Gln Tyr Asp Gln Ile Pro Ile 50 55 60	192
gaa att tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240

This page is not part of the pamphlet!

WO 01-35316 3/5

Date: 17 may 2001

Destination: Agent

-40-

						gga Gly										Thr		288
1:						agt Ser												336
	cca Pro	gga Gly	atg Met 115	gat Asp	ggc	cca Pro	aga Arg	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu		384
						aca Thr												432
						999 Gly 150												480
						aat Asn												528
•	agg Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly		576
						gga Gly												624
						tat Tyr												672
						acc Thr 230												720
						aat Asn												768
						agc Ser												816
						gtt Val												864
						ata Ile											:	912
						agg Arg 310											!	960

-41-

Glr	aaa Lys	gaa Glu	cct	cca Pro 325	Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa _Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aar Lys	gac Asp	agt Ser 350	tgg Trp	acw Xaa	1056
			ata Ile					Gly								1104
	tac Tyr 370	Pro	Gly aaa													1116
<21 <21	0 > 2 1 > 1 2 > D 3 > H	116 NA	Imm	unod:	ific	iency	y Vi:	rus	(HIV)	,						
<22	1> C 2> (0)	. (29° rotea													
<22	1 > Cl 2 > (: 3 > Pe	298)	(i			zers <i>e</i>	o Tra	angci	ei me e							
									TPC	ise						
cct	0> 24 cag Gln	atc	act Thr	ctt	tgg	caa	cga	ccc	ata	gtc	aca Thr	ata Ile	aag Lys	ata Ile 15	gly aga	48
cct Pro 1	cag Gln caa	atc Ile cta	act	ctt Leu 5 gaa	tgg Trp gct	caa Gln cta	cga Arg	ccc Pro	ata Ile 10	gtc Val gga	Thr	Ile gat	Lys gat	Ile 15 aca	Gly gta	4 8 96
cct Pro 1 999 Gly	cag Gln caa Gln gaa	atc Ile cta Leu	act Thr aag Lys	ctt Leu 5 gaa Glu aat	tgg Trp gct Ala	caa Gln cta Leu	cga Arg ata Ile	ccc Pro gat Asp 25	ata Ile 10 aca Thr	gtc Val gga Gly	Thr gca Ala cca	Ile gat Asp	Lys gat Asp 30 tta	11e 15 aca Thr	gta Val	
cct Pro 1 ggg Gly tta Leu	cag Gln caa Gln gaa Glu	atc Ile cta Leu gac Asp 35	act Thr aag Lys 20	ctt Leu 5 gaa Glu aat Asn	tgg Trp gct Ala ttg Leu	caa Gln cta Leu cca Pro	cga Arg ata Ile gga Gly 40	ccc Pro gat Asp 25 aga Arg	ata Ile 10 aca Thr tgg Trp	gtc Val gga Gly aaa Lys	Thr gca Ala cca Pro	gat Asp aaa Lys 45	gat Asp 30 tta Leu	Ile 15 aca Thr ata Ile	gta val ggg Gly ata	96
cct Pro 1 ggg Gly tta Leu gga Gly	cag Gln caa Gln gaa Glu att Ile 50 att	cta Leu gac Asp 35 gga Gly	act Thr aag Lys 20 ata Ile	ctt Leu 5 gaa Glu aat Asn ttt Phe	tgg Trp gct Ala ttg Leu gtc Val	caa Gln cta Leu cca Pro aga Arg 55	cga Arg ata Ile gga Gly 40 gtg Val	gat Asp 25 aga Arg aaa Lys	ata Ile 10 aca Thr tgg Trp cag Gln	gtc Val gga Gly aaa Lys tat Tyr	Thr gca Ala cca Pro gat Asp 60	gat Asp aaa Lys 45 cag Gln	gat Asp 30 tta Leu ata Ile	Ile 15 aca Thr ata Ile ccc Pro	Gly gta Val ggg Gly ata Ile	96 144
cct Pro 1 ggg Gly tta Leu gga Gly gaa Glu 65	cag Gln caa Gln gaa Glu att Ile 50 att Ile	cta Leu gac Asp 35 gga Gly tgt Cys	act Thr aag Lys 20 ata Ile ggt Gly	ctt Leu 5 gaa Glu aat Asn ttt Phe cat His	tgg Trp gct Ala ttg Leu gtc Val aaa Lys 70	caa Gln cta Leu cca Pro aga Arg 55 gtt Val	cga Arg ata Ile gga Gly 40 gtg Val ata Ile	gat Asp 25 aga Arg aaa Lys	ata Ile 10 aca Thr tgg Trp cag Gln aca Thr	gtc Val gga Gly aaa Lys tat Tyr gta Val 75	Thr gca Ala cca Pro gat Asp 60 tta Leu cag	gat Asp aaa Lys 45 cag Gln gta Val	gat Asp 30 tta Leu ata Ile gga Gly	Ile 15 aca Thr ata Ile ccc Pro	Gly gta Val ggg Gly ata Ile aca Thr 80 act	96 144 192

-42-

cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aga Arg	gtt Val 120	Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
			gca Ala													432
			aaa Lys													480
			aag Lys													528
			aat Asn 180													576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	gga Gly	tta Leu	aaa Lys 200	aag Lys	aac Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
			gat Asp													672
			gcg Ala													720
			cag Gln													768
			caa Gln 260													816
			gaa Glu													864
			tta Leu													912
aga Arg 305	saa Xaa	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	ttc Phe	aca Thr 315	cca Pro	gac Asp	caa Gln	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
		Thr	gta Val 340													1056

-43-

gty aat gac ata cag aaa tta gtk gga aaa ttg aat tgg gca agt caa Xaa Asn Asp Ile Gln Lys Leu Xaa Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tac cca ggg Ile Tyr Pro Gly 370	1116
<210> 25 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 25 cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aaa ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta aag gaa gct cta cta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg agt ttg cca gga aaa tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Ser Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta tcc atg Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Ser Met 50 55 60	192
gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga tct aca Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Ser Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ytg ttg act cag ctt ggg tgc act Pro Val Asn Ile Ile Gly Arg Asn Xaa Leu Thr Gln Leu Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gar ggg Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135 140	432

-44-

aaa Lys 145	Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
					gac Asp											528
					aaa Lys											576
					Gly 999											624
					tat Tyr											672
					acc Thr 230											720
					aat Asn											768
					agc Ser											816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tac Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
					ata Ile											912
					agg Arg 310											960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctc Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
					cct Pro											1056
					aag Lys											1104
	tac Tyr 370															1116

-45-

```
<210> 26
  <211> 1116
  <212> DNA
  <213> Human Immunodificiency Virus (HIV)
~~<220>
 <221> CDS
  <222> (0) ... (297)
  <223> HIV Protease
 <221> CDS
 <222> (298) ... (1116)
 <223> Portion of HIV Reverse Transcriptase
 <400> 26
 cct cag atc act ctt tgg caa cga ccc atc gtc gaa ata aag gta ggg
Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Glu Ile Lys Val Gly
                                                                                          48
 ggg caa cta ata gaa gct cta tta gat aca gga gca gat gat aca gta
                                                                                          96
 Gly Gln Leu Ile Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
 tta gaa gaa ata aat tta cca gga aga tgg aaa cca aga atg ata ggg
                                                                                         144
 Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Arg Met Ile Gly
 gga att gga ggt ttt gtc aaa gta aga cag tat gat cag gta cct atc
                                                                                         192
 Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Val Pro Ile
 gaa atc tgt gga cat aaa gtt ata agt aca gta tta gta gga cct aca
Glu Ile Cys Gly His Lys Val Ile Ser Thr Val Leu Val Gly Pro Thr
                                                                                         240
 cct gcc aac ata att gga aga aat ctg atg act cag att ggt tgc act
Pro Ala Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr
                                                                                         288
 tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aaa
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
                                                                                         336
                100
 cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa
Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu
                                                                                         384
 aaa ata aaa gca tta gta gaa att tgt aca gaa ytg gaa gag gaa ggg
                                                                                         432
 Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Xaa Glu Glu Glu Gly
                               135
 aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt
                                                                                        480
 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe
                          150
gcc ata aag aag aaa nnn agt ggt aga tgg aga aaa ata gta gat ttt
                                                                                        528
Ala Ile Lys Lys Lys Xaa Ser Gly Arg Trp Arg Lys Ile Val Asp Phe
```

-46-

ag: Ar	a gaa g Glu	a ctt 1 Lei	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gat Asp 185	Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Let	gga Gly	576
ata Ile	a cca e Pro	Cat His 195	Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	Lys	aac Asn	aag Lys	tca Ser	yta Val 205	Thr	att	ctg Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	Lys	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aat Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	cag Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gat Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	gag Glu	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	car Gln	cat His	ctg Leu	tta Leu	arg Xaa 310	tgg Trp	gga Gly	ttt Phe	ttc Phe	aca Thr 315	cca Pro	gaa Glu	caa Gln	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	ccm Xaa 325	ttc Phe	cak Xaa	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cay His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cas Xaa	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gat Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile	tac Tyr 370	cca Pro	ggg ggg													1116

<210> 27 <211> 1113 <212> DNA <213> Human Immunodificiency Virus (HIV)

-47-

<221> CDS <222> (0)...(297) <223> HIV Protease <221> CDS <222> (298)...(1116) <223> Portion of HIV Reverse Transcriptase cct cag atc act ctt tgg caa cga ccc atc gtc gaa ata aag gta ggg 48 Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Glu Ile Lys Val Gly ggg caa cta ata gaa gct cta tta gat aca gga gca gat gat aca gta 9.6 Gly Gln Leu Ile Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val tta gaa gaa ata aat tta cca gga aga tgg aaa cca aga atg ata ggg Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Arg Met Ile Gly 144 gga att gga ggt ttt gtc aaa gta aga cag tat gat cag gta cct atc 192 Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Val Pro Ile gaa atc tgt gga cat aaa gtt ata agt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Val Ile Ser Thr Val Leu Val Gly Pro Thr 240 cct gcc aac ata att gga aga aat ctg atg act cag att ggt tgc act Pro Ala Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr 288 90 tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aaa 336 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 110 cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gaa 384 Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 aaa ata aaa gca tta gta gaa att tgt aca gaa ytg gaa gag gaa ggg 432 Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Xaa Glu Glu Glu Gly aaa att tca aaa att ggg cct gaa aat cca tac aat act cca ata ttt 480 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe 150 155 528 Ala Ile Lys Lys Lys Ser Gly Arg Trp Arg Lys Ile Val Asp Phe Arg gaa ctt aat aag aga act caa gat ttc tgg gaa gtt caa tta gga ata Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile 576 CCa cat ccc gca ggg tta aaa aag aac aag tca gta aca att ctg gat Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Ile Leu Asp 624

-48-

Ş	gtg /al	ggt Gly 210	gat Asp	gca Ala	tat Tyr	ttt Phe	tca Ser 215	gtt Val	ccc Pro	tta Leu	gat Asp	aag Lys 220	gaa Glu	ttc Phe	agg Arg	aag Lys	672
ı	at Yr 25	act Thr	gca Ala	ttt Phe`	acc Thr	ata Ile 230	cct Pro	agt Ser	ata Ile	aat Asn	aat Asn 235	gag Glu	aca Thr	cca Pro	Gly 999	att Ile 240	720
A	iga Arg	tat Tyr	cag Gln	tac Tyr	aat Asn 245	gtg Val	ctt Leu	cca Pro	cag Gln	gga Gly 250	tgg Trp	aaa Lys	gga Gly	tca Ser	cca Pro 255	gca Ala	768
I	ata [le	ttc Phe	caa Gln	agt Ser 260	agc Ser	atg Met	aca Thr	aaa Lys	atc Ile 265	tta Leu	gag Glu	cct Pro	ttt Phe	aga Arg 270	aaa Lys	caa Gln	816
. P	at Asn	cca Pro	gac Asp 275	ata Ile	gtt Val	atc Ile	tat Tyr	cag Gln 280	tac Tyr	gtg Val	gat Asp	gat Asp	ttg Leu 285	tat Tyr	gta Val	gga Gly	864
t	ct Ser	gat Asp 290	tta Leu	gaa Glu	ata Ile	Gly 999	gag Glu 295	cat His	aga Arg	aca Thr	aaa Lys	ata Ile 300	gag Glu	gaa Glu	ctg Leu	aga Arg	912
C	ar 31n 305	cat His	ctg Leu	tta Leu	arg Xaa	tgg Trp 310	gga Gly	ttt Phe	ttc Phe	aca Thr	cca Pro 315	gaa Glu	caa Gln	aaa Lys	cat His	cag Gln 320	960
ī	aaa Lys	gaa Glu	cct Pro	ccm Xaa	ttc Phe 325	cak Xaa	tgg Trp	atg Met	ggt Gly	tat Tyr 330	gaa Glu	ctc Leu	cay His	cct Pro	gat Asp 335	aaa Lys	1008
t	L. G. B.	aca Thr	gta Val	cas Xaa 340	cct Pro	ata Ile	gtg Val	ctg Leu	cca Pro 345	gaa Glu	aaa Lys	gat Asp	agc Ser	tgg Trp 350	act Thr	gtc Val	1056
ī	aat Asn	gac Asp	ata Ile 355	cag Gln	aag Lys	tta Leu	gtg Val	gga Gly 360	aaa Lys	ttg Leu	aat Asn	tgg Trp	gca Ala 365	agt Ser	cag Gln	att Ile	1104
		cca Pro 370															1113
	212	0 > 28 L > 11 2 > Di 3 > Hu	L16 VA	Immi	ınod:	lfic	Lency	y Vi:	rus	(HIV))						
<	:22	L> CI 2> (())	. (29° rotea	7) ase												
	:22	l> CI 2> (2 3> Po	298)	(: on o	1116) E HIV) V Rev	vers	e Tra	ansci	ripta	ase						

-49-

	< 40	0 > 2	28														
(cct	caa	ato								Val					ggg Gly	4.8
					Glu	a gct ı Ala				Thr					Thr	gta Val	96
t I	ta Leu	gaa Glu	gaa Glu 35	. Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	Arg	tgg Trp	aag Lys	cca Pro	aaa Lys 45	Met	ata Ile	gtg Val	144
ć	ga Sly	att Ile 50	Gly	ggt Gly	ttt Phe	agc Ser	aaa Lys 55	gta Val	aga Arg	caa Gln	tat Tyr	gat Asp 60	Gln	ata Ile	ccc Pro	ata Ile	192
						aaa Lys 70											240
						gga Gly											288
						agt Ser											336
						cca Pro											384
a L	aa ys	ata Ile 130	aaa Lys	gca Ala	tta Leu	ata Ile	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	gaa Glu	gma Xaa	gga Gly	432
L						999 Gly 150											480
9	cc la	ata Ile	aag Lys	aaa Lys	aar Lys 165	aac Asn	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gac Asp 175	ttc Phe	528
a:	gg rg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	Gln	gac Asp 185	Phe	Trp	Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
a!	ta le	cca Pro	cat His 195	cct Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
ga As	sp '	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	att Ile	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aa Ly 22	/S	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	Thr	aat Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720

PCT/US00/30863

-50-

att Ile	aga Arg	tat Tyr	cag Gln	tat Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	ctt Leu	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gag Glu 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gag Glu	cta Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	tgg Trp	aag Lys 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	ccc Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gka Xaa 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	caa Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
	tat Tyr 370															1116
<212 <212	0> 29 L> 11 2> DN 3> Hu	.16 IA	Immu	modi	.fici	.ency	, Vir	rus ((HIV)			•				
<220)>															
<221 <221	L> CI 2> (0 3> HI)														
<222	l> CI 2> (2 3> Po	98).				erse	: Tra	nscr	ripta	ıse						
)> 29															
cct Pro 1	cag Gln	atc Ile	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	Pro	Ile 10	gtc Val	aca Thr	ata Ile	aag Lys	Ile 15	Gly aga	48
ggg Gly	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	ata Ile	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96

						•										
tta Leu	gaa Glu	gaa Glu 35	atg Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	ata Ile	ata Ile	Gly aaa	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ctt Leu	gtc Val	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	ccc Pro	ata Ile	192
gaa Glu .65	atc Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gtt Val	ata Ile	ggt Gly	aca Thr	gtw Xaa 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gcc Ala	aác Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	ggg Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gag Glu	aag Lys	gag Glu	gga Gly	432
aag Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aag Lys 165	aac Asn	agt Ser	act Thr	agg Arg	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aac Asn	aaa Lys	tca Ser	gca Ala 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggc Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	Leu	gac Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acy Xaa 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gaa Glu	aca Thr	cca Pro	ggg Gly 240	720
tar Xaa	ata Ile	tca Ser	gtg Val	tac Tyr 245	aat Asn	gtr Xaa	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cma Xaa	768
gca Ala	ata Ile	ttc Phe	maa Xaa 260	agt Ser	agc Ser	atg Met	aca Thr	aga Arg 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816

caa a Gln A	aat co Asn Pr 27	O GIU	ata Ile	gtt Val	atc Ile	tat Tyr 280	GIL	tac Tyr	gtg Val	gat Asp	ga: Asj 28) Le	g ta ı Ty:	t gta r Val	864
GIY 3	ct ga Ser As 190	c tta p Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	[Va]	a gag L Glu	g gaa 1 Glu	a ctg ı Leu	912
aga c Arg G 305	aa ca Un Hi	t ctg s Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	ttc Phe	aca Thr 315	cca Pro	gac	caa Glm	aaa Lys	cat His 320	960
cag a Gln L	aa gaa ys Gli	a ccc a Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	ect Pro	Asp	1008
aaa to Lys T	gg aca rp Thi	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act	1056
gtc aa Val As	at gad sn Asp 355	TIE	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att ta Ile Ty 37	r Ala	gl ^A aaa													1116
<210><211><212><213>	1116 DNA	Immu	nodi:	fici	encv	Vir	11 e (HTW)							
<220><221><222><223>	CDS (0)	. (297))					,							
<221> (<222> <223>)	(298)	(11 on of	L16) HIV	Reve	erse	Trai	nscr:	ipta	se						
<400> 3 cct cas Pro Glr 1	a atc	act o	tt t eu T 5	tb e	aa c ln A	ga d rg I	ecc o	cty (Kaa 1	gtc a Val 1	aca a Thr :	ata Ile	aag Lys	ata Ile 15	gly aaa	48
ggg caa Gly Glr	a cta 1 Leu	aag g Lys G 20	raa g lu A	ct y la X	ta t aa L	ta g eu A	gat a Asp 1 25	ca g	gga g Bly A	gca g Nla A	gat Asp	gat Asp 1	aca Thr	gta Val	96
tta gaa Leu Glu	gaa Glu 35	atg a Met S	gc t er L	ta c eu P	ro G	ga a ly A 40	ga t Ig I	gg a	aa c ys P	ca a ro I	aa a ys l 45	atg a	ata Ile	gly ggg	144
gga att Gly Ile 50	GIY.	ggk t Kaa P	tt a he I	Te Ti	aa gi ys Va 55	tg a al X	gm c aa G	ag t ln T	уr А	at c sp G 60	ag a	ata d [le]	etc a Leu :	ata Ile	192

Gl	a at u Xa 5	y tg a Cy:	t gga s Gly	a cat y His	t aaa s Lys 70	s Ala	ata 116	a ggt e Gly	aca Thi	a gti r Xaa 75	a Lei	a ata 1 Ile	a gga e Gly	a cct y Pro	aca Thr 80	240
cc Pr	t gto	c aad l Asi	c ata n Ile	a att	∍ Glչ	a aga / Arc	aat JAsr	ctg Leu	tto Lev 90	ı Thi	cac Glr	g att	ggt Gly	tgo Cys 95	act Thr	288
tt Le	a aat u Asi	tti n Phe	ccc Pro) Ile	agt Ser	Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Lev	aaa Lys	336
Pr	a gga o Gly	A ato Met	: Asp	ggc Gly	cca Pro	aaa Lys	gtc Val 120	Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	Thr	gaa Glu	gaa Glu	384
aa Ly	a ata s Ile 130	: Lys	a gca s Ala	tta Leu	gta Val	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	Glu	aag Lys	gaa Glu	ggr Xaa	432
aa: Ly: 14!	a att s Ile	aca Thr	aaa Lys	att	999 Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gco	ata a Ile	aag Lys	aag Lys	aaa Lys 165	Asn	agt Ser	gat Asp	aaa Lys	tgg Trp 170	Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	a gaa g Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cca Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	cag Gln	aaa Lys	aag Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gta Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
gtt Val	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gta Val	ctc Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttt Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	agg Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gaa Glu	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	Ile	999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912

-54-

											•					
ag Ar 30	g Gl	a cat n His	t cto s Lev	g tto 1 Lem	agg Arg 310	Trp	gga Gly	ttc Phe	tac Tyr	aca Thr 315	Pro	a gad Asp	Lys	a aaa S Lys	cat His 320	960
Ca Gl	g aaa n Lys	a gaa s Gli	a cct ı Pro	Pro 325	Phe	ctt Leu	tgg Trp	g atg Met	ggt Gly 330	Tyr	gaa Glu	cto Leu	cat His	cct Pro	gat Asp	1008
aaa Lys	a tgg s Trp	g aca Thi	a gta Val 340	. Glr	cct Pro	ata Ile	gtg Val	Leu 345	cca Pro	gaa Glu	aaa Lys	gad	e ago Ser 350	Trp	act Thr	1056
gto Val	aat L Asr	gac Asp 355) Ile	cag Glr	aag Lys	tta Leu	gta Val 360	Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	Ala	agt Ser	cag Gln	1104
		Āla	gga Gly										•	•		1116
<21 <21	.0> 3 .1> 1 .2> D .3> H	.117 NA	Imm	unod	ific	ienc	y Vi	rus	(HIV))					0 Y 0	
<22	1> C 2> (0)	.(29 rote													
<22		298)	(; on o:			verse	e Tra	ansci	ripta	ase						
cct	0> 3 cag Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	ctc Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	Gly aaa	48
gly ggg	caa Gln	tta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
cta Leu	gaa Glu	gac Asp 35	gtg Val	cat His	ttg Leu	cca Pro	gga Gly 40	aaa Lys	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	gly aaa	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	gag Glu	gta Val	ccc Pro	ata Ile	192
gaa Glu 65	ctc Leu	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
ccc Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	wtg Xaa 90	act Thr	caa Gln	ctt Leu	ggy ggg	tgc Cys 95	act Thr	288

) Ile					Thr					Lev	aag Lys	336
			Asp					. Lys					Thr		gaa Glu	384
aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	Glu	aag Lys	gaa Glu	Gly	432
	Ile					Pro		aat Asn			Asn					480
gyc Xaa	ata Ile	aag Lys	aaa Lys	aaa Lys 165	Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
								gac Asp 185								576
								aag Lys								624
								gtt Val								672
								agt Ser								720
att Ile	aga Arg	tac Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
								aaa Lys 265								816
caa Gln	aac Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tcy Xaa 290	gac Asp	tta Leu	gaa Glu	ata Ile	gga Gly 295	cag Gln	cat His	agr Xaa	rca Xaa	aaa Lys 300	ata Ile	gaa Glu	gaa Glu	ctg Leu	912
								ttt Phe								960
car Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttt Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008

-56-

aaa Lys	tgg Trp	aca Thr	gtg Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ant Xaa	aca Thr	gaa Glu	gtt Val	agt Ser 360	ggg ggg	aaa Lys	att Ile	gaa Glu	ttg Leu 365	ggc Gly	aag Lys	tca Ser	1104
	tta Leu 370			g												1117
<21 <21	0 > 3: 1 > 1: 2 > Di 3 > Hi	116 NA	Imm	unod:	ific:	iency	y Vi:	rus ((HIV)						
<22	0> 1> Cl 2> ((3> H)	0)														
<22	1> Cl 2> (; 3> Po	298)				/erse	e Tra	ansci	cipta	ase						
cct	0> 3: caa Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	cty Xaa 10	gtc Val	gca Ala	ata Ile	agg Arg	ata Ile 15	gly ggg	48
Gly 999	caa Gln	cta Leu	aag Lys 20	gaa Glu	gcc Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta Leu	gaa Glu	gac Asp 35	atg Met	gag Glu	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aag Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly 999	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aam Xaa	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	ctt Leu	gta Val	192
gaa Glu 65	atc Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	gta Val	ggt Gly	aca Thr	gta Val 75	tta Leu	ata Ile	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ttg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	ggc Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Il e	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gag Glu	384

	aa Ly	a at s II	.е ь	aa g ys <i>l</i>	gca Ala	tta Leu	gta Val	ga L Gl: 13	u Il	c tg e Cy	t ac s Th	a ga r Gl	a tt u Le 14	u Gl	a aa u Ly	g ga s Gl	a gga u Gly	432
-	aa Ly 14	S 11	t t e S	ca a er I	aaa Lys	att Ile	999 Gly 150	Pro	t ga o Gli	a aa u Ası	t cc n Pr	a tac o Ty: 15!	r As	t ac n Th	t cc r Pr	a gt o Va	a ttt l Phe 160	480
	gc Al	t at a Il	a a e L	ag a ys I	aa .ys	aaa Lys 165	gac Asp	agt Sei	act	t aaa c Lys	a tgg s Trj 170	o Arg	a aa g Ly:	a tt: s Le:	a gta	a gat l Asj 17!	t ttc p Phe	528
	ag Ar	a ga g Gl	a ci u Le	eu A	at sn 80	aaa Lys	aga Arg	act Thr	caa Glr	a gad n Asp 189	Phe	tgg Trp	g gaa o Gli	a gti u Val	caa l Gli 190	ı Let	a gga ı Gly	576
	ata Ile	a cc e Pr	a ca o Hi 19	s P	cc ro	gca Ala	Gly 999	tta Leu	aaa Lys 200	Lys	aaa Lys	a aaa ; Lys	tco Sei	gtg Val	l Thi	a gta Val	a ctg Leu	624
	gat As <u>r</u>	ya: Va: 21	I GI	jt g .y A	at sp	gca Ala	tat Tyr	ttt Phe 215	Ser	gtt Val	Pro	tta Leu	gat Asp 220	Lys	gac Asp	ttt Phe	aga Arg	672
	aag Lys 225	Ty	ac Th	t g	ca i la i	ttt Phe	acc Thr 230	aya Xaa	cct	sgt Xaa	ata Ile	aac Asn 235	Asr	gag Glu	, aca Thr	cca Pro	999 Gly 240	720
	att Ile	aga Arg	ta y Ty	t car	rn 1	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	Gly	tgg	aaa Lys	gga Gly	Ser 255		768
	gca Ala	ata Ile	tt Ph	t ca e G] 26	Ln S	agc Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
	caa Gln	aat Asn	27	o As	k q	vta (aa	gtt Val	wtc Xaa	tat Tyr 280	caa Gln	twc Xaa	ata Ile	gat Asp	gat Asp 285	ctg Leu	tat Tyr	gta Val	864
	ggc Gly	tct Ser 290	AS	tt Le	a g u G	lu lu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
	aga Arg 305	cag Gln	cat His	ct Le	g t u T	Lb :	aag Lys 310	tgg Trp	ggg Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
	cag Gln	aaa Lys	gaa Glu	cc Pr	O P	ca t ro 1 25	ttt Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
	aaa Lys	tgg Trp	aca Thr	Va.	I G.	ag d ln I	ect (Pro	ata Ile	atg Met	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
	gtc Val	aat Asn	gac Asp 355	TT	a ca e Gi	ag a ln I	ar i	Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104

WO 01/35316 PCT/US00/30863

-58-

att tac cca ggg Ile Tyr Pro Gly 370	1116
<210> 33 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<400> 33 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1 5 10 15	48
ggg caa cta aag gaa gct cta tta kat aca gga gca gat gat aca gtm Gly Gln Leu Lys Glu Ala Leu Leu Xaa Thr Gly Ala Asp Asp Thr Xaa 20 25 30	96
tta gaa gac atg act ttg cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Asp Met Thr Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aaa cag tat gag gag ata ccc ata Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Glu Glu Ile Pro Ile 50 55 60	192
gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ttg ttg act cag att ggt tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aaa Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca ttw gta gaa att tgt gca gaa ctg gaa aag gaa ggg Lys Ile Lys Ala Xaa Val Glu Ile Cys Ala Glu Leu Glu Lys Glu Gly 130 135 140	432
aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 145 150 155 160	480

-59-

gc: Ala	c ata	a aaq e Lys	g aaa s Lys	a aaa Eys 165	Asp	ggt Gly	act Thr	aaa Lys	tgg Trp	Arg	aag J Lys	gta Val	aca Thr	gat Asp 175	ttt Phe	528
aga Arg	a gaa g Glu	a ctt 1 Leu	aat 1 Asr 180	Lys	agg Arg	ach Xaa	caa Gln	gac Asp 185	Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Lev	gga Gly	576
ata Ile	cca Pro	cat His	Pro	tca Ser	Gly	tta Leu	aaa Lys 200	Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	. Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	Asn	gcg Ala	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	atg Met	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gat Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	ggt Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttt Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
_	tat Tyr 370															1116

<210> 34 <211> 1119 -60-

	.2> D .3> H		Imm	unod	ific	ienc	y Vi	rus	(HIV	·)						
<22	1> C 2> (DS 0)														
<22		298)				vers	e Tr	ansc	ript	ase						
cct		atc													Gly 999	4.6
								gac Asp 25								96
								aga Arg								144
								aaa Lys								192
gam Xaa 65	atc Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	aca Thr	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
								atg Met								288
								gaa Glu 105								336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aga Arg	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
Lys	Ile	Lys	Āla	Leu	V al	Ğlu	Ile	tgt Cys	Thr	Ğlu	Leu	Ğlu				432
								aat Asn								480
								aaa Lys								528
aga Arg	gaa Glu	ctt Leu	aac Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576

-61-

ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	Gly 999	tta Leu	cca Pro 200	aag Lys	aac Asn	aaa Lys	tca Ser	gta Val 205	acg Thr	gta Val	ctg Leu	624
gat _Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	cct Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tac Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agg Arg	tat Tyr	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
act Thr	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cct Pro	atg Met 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aga Arg	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gac Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gag Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	gcg Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	gaa Glu	cat His	ctg Leu	tgg Trp	aag Lys 310	tgg Trp	ggt Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	cat His	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aaa Lys	tta Leu	gtg Val 360	ggr Xaa	aaa Lys	att Ile	gaa Glu	ttt Phe 365	ggg Gly	cga Arg	gtc Val	1104
Arg		amc Xaa														1119
<210 <211		.15														

<212> DNA <213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS <222> (0)...(297) <223> HIV Protease

-62-

<221> CDS <222> (298)(1115) <223> Portion of HIV Reverse Transcriptase																
cct		ato								Val					gly	48
gly aas	caa Gln	tta Leu	aag Lys 20	Glu	gct Ala	cta Leu	cta Leu	gat Asp 25	Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gta Val	96
tta Leu	gaa Glu	gac Asp 35	Met	aat Asn	tta Leu	cca Pro	gga Gly 40	Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly 999	144
		Gly										cag Gln				192
gaa Glu 65	Ile	tgt Cys	gly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
												gta Val				336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
												gaa Glu				432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	gga Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aag Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gaa Glu	gtc Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	Gly 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	tta Leu	624
gat Asp	gtg Val 210	gga Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672

WO 01/35316 PCT/US00/30863

-63-

aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aag Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtc Val	ata Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
Gly 999	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gga Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cac His	ttg Leu	ttg Leu	maa Xaa 310	tgg Trp	gga Gly	ttc Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aag Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	ccc Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	kaa Xaa	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	ctg Leu	1056
tca Ser	atg Met	aca Thr 355	tac Tyr	aga Arg	aat Asn	tag *	tgg Trp	gaa Glu 360	agt Ser	tga *	att Ile	gly ggg	caa Gln	gtc Val 365	aaa Lys	1104
	atg Met	cng Xaa	3 3													1115
<21 <21	0 > 36 1 > 1: 2 > DI 3 > Hi	116	Imm	ınodi	ifici	iency	y Vi:	rus	(HIV)							
<22	1> CI 2> ((OS O) IV Pi	-	-												
<22	1> CI 2> (2 3> Po	OS 298) ortic	(1 on of	116) HIV	/ / Rev	/erse	e Tra	ansci	ripta	ase						
cct	0> 36 cag Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	cca Pro	gtc Val 10	gtc Val	aca Thr	ata Ile	aag Lys	gta Val 15	gly aaa	48

-64-

,	cad	cta	aaα	gaa	act	cta	tta	gat	aca	gga	gca	gat	gat	aca	gta		96
Gly	Gln	Leu	Lys 20	Glu	Ala	Leu	Leu	Asp 25	THE	GIY	Ala	Asp	30	****	• • • •		
tta Leu	gaa Glu	gaa Glu -35	atg Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	aga Arg	tġg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly 393		144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	rtc Xaa	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	caa Gln	ata Ile	ccc Pro	ata Ile		192
gaa Glu 65	atc Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	aca Thr	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	eet Pro	aca Thr 80		240
cct Pro	gyc Xaa	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	ggg Gly	tgc Cys 95	act Thr		288
tta Leu	aat Asn	ttt Phe	cct Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys		336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ctg Leu 125	aca Thr	gaa Glu	gaa Glu		384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	gca Ala	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	Gly aaa		432
aag Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	ccy Xaa	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aay Asn	act Thr	cca Pro	gta Val	ttt Phe 160		480
gcc Ala	ata Ile	aag Lys	aaa Lys	aar Lys 165	aac Asn	agt Ser	act Thr	ara Xaa	tgg Trp 170	aga Arg	aaa Lys	kta Xaa	gta Val	gat Asp 175	ttc Phe		528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gat Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	-	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	gly ggg	cta Leu	aag Lys 200	ьys	aaa Lys	aaa Lys	tca Ser	gta Val 205	TITE	gta Val	ctg Leu		624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	ttg Leu	gat Asp 220	GIU	gac Asp	ttc Phe	aga Arg		672
aag Lys 225	Tyr	aca Thr	gcc Ala	ttt Phe	acc Thr 230	Tyr	act Thr	ggt Gly	tcc Ser	aac Asn 235	ASII	gag Glu	aca Thr	cca Pro	999 Gly 240		720
att Ile	aga Arg	tat Tyr	car Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	GIY	tgg Trp	aaa Lys	gga Gly	tca Ser 255	Pro		768

WO 01/35316 PCT/US00/30863

-65-

gca ata tto Ala Ile Phe				Val Leu					816
caa aat cca Gln Asn Pro 275	Asp Ile					p Leu			864
gga tct gac Gly Ser Asp 290									912
aga caa cat Arg Gln His 305					Pro As				960
cag aaa gaa Gln Lys Glu									1008
aaa tgg aca Lys Trp Thr	gta cag Val Gln 340	cct ata Pro Ile	gtg ctg Val Leu 345	cca gaa Pro Glu	aaa ga Lys As	c agc p Ser 350	tgg Trp	act Thr	1056
gtt aat gac Val Asn Asp 355	ata cag Ile Gln	aaa tta Lys Leu	gtg gga Val Gly 360	aaa ttg Lys Leu	aat tg Asn Tr 36	p Ala	agt Ser	cag Gln	1104
att tac cca Ile Tyr Pro 370									1116
<210> 37 <211> 1116 <212> DNA <213> Human	Immunodi	.ficiency	<i>y</i> Virus	(HIV)				٠	
<220> <221> CDS <222> (0) <223> HIV P									
<221> CDS <222> (298). <223> Portic		Reverse	: Transcr	riptase					
<400> 37 cct cag atc Pro Gln Ile 1									48
ggg caa cta Gly Gln Leu									96
tta gaa gac Leu Glu Asp 35						s Met			144

G1	y II	t gg e Gl 0	ga gg .y Gl	t tt y Ph	t at e Il	c aaa e Ly: 5:	s Va	a aga l Arg	a caq g Gli	g ta n Ty	t ga r As 6	p Gl:	g gt n Va	a cc l Pr	c ata o Ile	192
ga Gl	u II	c tg e Cy	t gg s Gl	a ca y Hi	t aa s Ly: 7	s Ala	ata a Ile	a ggi e Gly	t aca y Thi	a gta r Val	l Le	a gta	a gg l Gl	a cc y Pro	t aca o Thr 80	240
cc Pr	t gt o Va	c aa l As	c at n Il	a at e Ile 8!	e Gly	a aga y Arg	a aat g Asr	cto Lei	g ato Met 90	Th	a caq r Gli	g cti n Lei	t gg u Gl	t tgi y Cys 95	t act s Thr	288
tt. Le	a aa u Asi	t tt n Ph	t cc e Pro 10	o Ile	e Ser	cct Pro	att	gaa Glu 105	ı Thr	gta Val	e cca l Pro	a gta o Val	a aa: l Ly: 110	s Lev	a aag 1 Lys	336
Pro	a gga o Gly	a ato Men 11	t Ası	ggc Gly	c cca / Pro	aaa Lys	gtt Val 120	. Lys	caa Gln	tgc Trp	g cca Pro	tto Lev 125	ı Thi	a gaa Glu	gaa Glu	384
aaa Lys	a ata s Ile 130	; гл.	a gca s Ala	a tta a Lev	a gta ı Val	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	ato Met	: Glu	aag Lys	g gaa Glu	ggg	432
aaa Lys 145	: Ile	tca Sei	a aaa Lys	att Ile	999 Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	g aaa Lys	aaa Lys 165	Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	Phe	528
agg Arg	gaa Glu	Ctt Leu	aat Asn 180	Lys	aaa Lys	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Leu	Gly 999	576
ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	gga Gly	tta Leu	aaa Lys 200	aag Lys	aat Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	Thr	aaa Lys 265	att Ile	tta Leu	gat Asp	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
cag Gln	aat Asn	cca Pro 275	gat Asp	ata Ile	gtt Val	Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864

-67-

gga tct gac tta gag ata ggg cag cat aga gca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu 290 295 300	912
aga gca cat ctg ttg aag tgg gga ttt acc acc cca gac aaa aaa cat Arg Ala His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His 305 310 315 320	960
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tac gca ggg Ile Tyr Ala Gly 370	1116
<210> 38	
<211> 1117 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1117) <223> Portion of HIV Reverse Transcriptase	
<400> 38 cct caa tca ctt ctt tgg caa cga ccc mtc gtc aca ata aag gta ggg	48
Pro Gln Ser Leu Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Val Gly 1 10 15	40
ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca ata Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Ile 20 25 30	96
tta gaa gac aya rat ttg cca ggg aga tgg aaa cca aaa ata ata ggg Leu Glu Asp Xaa Xaa Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly 35 40 45	144
gga att gga ggt ttt atc aga gta aga cag tat gat cag gta ccc ata Gly Ile Gly Gly Phe Ile Arg Val Arg Gln Tyr Asp Gln Val Pro Ile 50 55 60	192
gaa atc tgt gga cat aaa gtt gta agt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Val Val Ser Thr Val Leu Val Gly Pro Thr 65 70 75 80	240

-68-

				J.1. I	ta a le I	85	эту	AL	g As	sn I	Leu	ме 9	0	nr (3ln	Il	e G	ly	Cys 95	Tl	nr	288
			·	1	cc a ro I 00	TE :	er.	PIC	> 11	Le (.05	Th	r Va	il F	,ro	۷a	l L ₃	/s :	Leu	ı L	/S	336
			11	5	at g sp G	IY E	10	пÀг	12	0	ys	GII	ı Tr	рP	ro	Le:	u Th	ır (Glu	. Gl	u	384
	7 ~	130	, <u>_</u> ,	s A.	a t	su v	aı	135	11	e C	ys	GIU	GI	u L	eu 40	Glı	ı Ly	s A	Asp	Gl	Y	432
ī	45		56.	<u>, 11</u>	a at	1	1 y 5 0	PIO	GI	u A	sn	Pro	15:	r A: 5	sn	Thr	Pr	0 V	'al	Ph	e 0	480
			цy.	y Ly	a aa s Ly 16	5 A	511	ser	Th	c Ti	ys	170	Arg	3 L	/S	Leu	Va.	l A 1	sp 75	Phe	2	528
	- 5 '			18		s Al	·9 .	IIII	GII	1 AS	35	Pne	Tr) G1	lu '	Val	Gl: 190	ı L	eu	Gl	7	576
			195		t gc > Al	a Gi	.y 1	Jeu	200	,	rs I	Lys	Lys	Se	r Y	Val 205	Thi	· Va	al	Leu	1	624
	2	10	CLY	noj	gc. Ala	д ту	2	15	ser	11	e i	ro	Leu	22	p (3lu	Asp	Pł	ne .	Arg	!	672
22	5	<i>y</i> =		VTC	tti Phe	23	0	тe	PIO	Se	r 1	.1e	Asn 235	As:	n G	lu	Thr	Pr	:O (Gly 240		720
		- 3	-1-		Tyr 245	, AS		a .	ьeu	PIC	2	50	GIÀ	Tr	o L	ys	Gly	Se 25	r 1	Pro		768
				260	agt Ser	50.	. 140		IIII	265	2 2	Te .	Leu	GI	1 P	ro	Phe 270	Ar	g I	ys		816
Glr	a aa a As		cca Pro 275	gac Asp	ata Ile	gto Val	at I		tat Tyr 280	caa Glr	ta T	at a yr N	atg let	gat Asp) A:	at sp:	ttg Leu	ta Ty:	t g r V	rta 'al		864
gga	Se 29		ac sp	tta Leu	gag Glu	ata Ile	99 G1 29	·y c	ag	cat His	aç Aı	ga a rg 1	ca hr	aaa Lys 300	I.	ta q le (gag Glu	gaa Glu	ı L	tg eu		912
aga Arg 305	ca Gl	g c n H	at is:	ctg Leu	tgg Trp	aag Lys 310	TT	р G 9 9	ly gg	ttt Phe	ta Ty	T T	ca hr 15	cca Pro	ga As	ac a sp 3	ara Kaa	aaa Lys	H	at is 20		960

-69-

	Glr	g aaa Lys	a gaa s Glu	a cct	Pro 325	Phe	ctt Leu	tgg Trp	g atg Met	ggt Gly 330	Tyr	gaa Glu	cto Lev	cat His	cct Pro 335	gac Asp	1008
	aaa Lys	tgg Tr	g aca Thr	gta Val	Glr	g cct n Pro	ata Ile	gtç Val	ctg Leu 345	Pro	gaa Glu	aag Lys	gac Asp	ago Ser 350	Trp	act Thr	1056
	gto Val	aat Asr	gac Asp 355	Ile	cag Glr	, aag Lys	tta Leu	gtg Val 360	Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	Ala	agt Ser	cag Gln	1104
			Xaa	agg Arg													1117
	<21 <21	0> 3 1> 1 2> E 3> H	128 NA	Imm	unod	ific	ienc	y Vi	rus	(HIV)						
	<22	1 > C 2 > (0)	.(29 rote	-												
	<22		298)	(: on o:			verse	e Tra	ansci	ript	ase						
	cct	0> 3 cag Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	cca Pro	ttc Phe 10	gtc Val	aca Thr	ata Ile	aaa Lys	ata Ile 15	glà aaa	48
	gly ggg	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	ata Ile	tta Leu	gac Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
	tta Leu	gaa Glu	gaa Glu 35	atg Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	gja aaa	144
	gga Gly	att Ile 50	Gly	ggt Gly	Phe	mtc Xaa	Lys	Val	aga Arg	cag Gln	Tyr	gat Asp 60	cag Gln	gta Val	ccc Pro	ata Ile	192
•	gaa Glu 65	atc Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gtt Val	atg Met	agt Ser	aca Thr	gta Val 75	tta Leu	ata Ile	gga Gly	cct Pro	aca Thr 80	240
]	cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	atg Met 90	act Thr	cag Gln	mtt Xaa	ggc Gly	tgc Cys 95	act Thr	288
1	tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gwa Xaa	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336

-70-

cca Pro	Gly	Met 115	Asp	ggc	Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	aat Asn	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gat Asp 185	ttc Phe	tgg Trp	gaa Glu	gtc Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	gly 999	tta Leu	aaa Lys 200	aag Lys	aac Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	raa Xaa	gat Asp	tca Ser	gra Xaa	672
agt Ser 225	aca Thr	ctg Leu	cat His	tta Leu	cca Pro 230	tac Tyr	cta Leu	gta Val	cgr Xaa	acc Thr 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
											gag Glu					816
											gat Asp					864
											aaa Lys 300					912
											cca Pro					960
											gaa Glu					1008
											aaa Lys					1056

-71-

tgt caa tga cat tac mag aaa gtt agt ggg gaa aat tgg aat ttg ggg Cys Gln * His Tyr Xaa Lys Val Ser Gly Glu Asn Trp Asn Leu Gly 355 360 365	1104
caa ggt cag att tat tgc cag ggg Gln Gly Gln Ile Tyr Cys Gln Gly 370 375	1128
<210> 40 <211> 1120 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1120) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 40 cct cag atc act ctt tgg caa cga ccc ctc gtt gca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ala Ile Lys Ile Gly 1</pre>	48
gga cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg agt ttg cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Ser Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga car tat gat cag ata ccm rta Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Xaa Xaa 50 55 60	192
gaa att tgc gga cat aaa gct gta ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ctg ttg act cag mtt ggt tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Xaa Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gaa act gta cca gtg aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135	432

aa Ly 14	S 11	t t .e S	ca a er L	aa a ys I	le G	gg c ly P 50	ct ga ro Gl	aa aa lu As	at co	ro 1	ac Tyr 155	Asr	ac 1 Th	t cc r Pr	a gt o Va	a ttt l Phe 160	:
gc Al	c at a Il	a aa e Ly	ag a ys L	ys L	aa ga ys A: 55	ac a sp S	gt ac er Th	t aa ir Ly	a to	cp A	ıga Irg	aaa Lys	tt: Le:	a gta	a gat l As _l 17!	t ttc p Phe	528
ag Ar	a ga g Gl	a ct u Le	eu A	at aa sn Ly 30	ag ag /s Ai	ga ad gg Tl	ct ca nr Gl	a ga n As 18	P P	c t ne T	gg 'rp	gaa Glu	gtt Val	caa l Glr 190	ı Let	a gga ı Gly	576
ata Ile	a cc e Pr	a ca o Hi 19	s Pı	t go O Al	a gg .a Gl	g tt y Le	a aa eu Ly 20	s Ly	g aa s Ly	a a 's L	aa ys	tca Ser	gta Val 205	. Thi	e gta Val	ctg Leu	624
gat As <u>r</u>	gte Va. 21	T GT	t ga y As	it go sp Al	a ta a Ty	t tt r Ph 21	ie Se	a gt r Va	t cc l Pr	t t	ta eu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agr Xaa	672
aag Lys 225	Ty	ac Th	t go r Al	a tt a Ph	t ac e Th 23	r Il	a cc e Pr	t ag o Se	t ac r Th	r A	ac sn 35	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	ta J Ty	t ca r Gl	g to n Se 24	r As	t gt n Va	g cti l Lei	t cca	a ca 5 Gl 25	n G	ga ly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	tt: Ph	c ca e Gl 26	n Se	t ag r Se	c at r Me	g aca t Thi	a aaa r Lys 265	s Il	c ct e Le	a eu	gaa Glu	cct Pro	ttt Phe 270	agg Arg	aaa Lys	816
caa Gln	aat Asn	Pro 27	O As	t at	a gti e Val	t at	c tat e Tyr 280	Glr	tac Ty	c at r Me	g et	gat Asp	gat Asp 285	cta Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	Asį	tt: Lei	a gaa 1 Gli	a ata 1 Ile	a ga e Gli 29!	a caç ı Glr	g cat His	aga Arg	a ac y Th	ır :	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat	cto Lei	g ttg 1 Lei	agg Arg 310	Tr	g ggg	ttt Phe	acc	ac Th	r 1	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	Pro	Pro 325	Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	' Ту	t c	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	GIn	Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	ga Gl	a a u I	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	хаа	aca Thr	aaa Lys	gtt Val	agt Ser 360	Gly 999	gaa Glu	aa Asi	t t n	ga (att Ile	999 Gly 365	sca Xaa	agt Ser	1104
cag Gln	att Ile	tat Tyr 370	tgg Trp	agg Arg	g												1120

-73-

```
<210> 41
    <211> 1059
    <212> DNA
    <213> Human Immunodificiency Virus (HIV)
 ~~<220>
   <221> CDS
   <222> (0)...(297)
   <223> HIV Protease
   <221> CDS
   <222> (298)...(1059)
   <223> Portion of HIV Reverse Transcriptase
   cct caa atc act ctt tgg cag cga ccc gtt gtc aca ata aac ata ggg
Pro Gln Ile Thr Leu Trp Gln Arg Pro Val Val Thr Ile Asn Ile Gly
                                                                                               48
  ggg caa cta aag gaa gct cta tta gac aca gga gca gat gat aca gta
Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
                                                                                               96
  tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg
Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
                                                                                              144
  gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata
Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
                                                                                              192
  gaa atc tgt gga cat aaa act ata ggt aca gta tta ata gga cct aca
Glu Ile Cys Gly His Lys Thr Ile Gly Thr Val Leu Ile Gly Pro Thr
                                                                                              240
 cct gtc aac ata att gga aga aat ctg ttg act cag att ggc tgc act
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr
                                                                                             288
 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag
                                                                                             336
 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
 cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa
                                                                                             384
 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
                                      120
 aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa ggg
                                                                                             432
 Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
      130
 aaa att tca aaa att ggg cct gaa aac ccg tac aat act cca gtc ttt
                                                                                             480
Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
gcc ata aag aaa aaa gat agt act aaa tgg aga aaa tta gta gat ttc
                                                                                            528
Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe
```

-74-

aga Arg	gaa Glu	ctt Leu	aac Asn 180	aag Lys	aaa Lys	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	att Ile	caa Gln 190	tta Leu	gga Gly	576
								aag Lys								624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	gtt Val	cct Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	acg Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gcc Ala	ata Ile	nnn Xaa	nnn Xaa 260	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa 265	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa 270	nnn Xaa	nnn Xaa	816
								caa Gln								864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gag Glu 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	aaa Lys	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gat Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttt Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	Thr	gta Val 340	cag Gln	cct Pro	ata Ile	Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val																1059

<210> 42 <211> 1053 <212> DNA <213> Human Immunodificiency Virus (HIV)

<220>

<221> CDS <222> (0)...(297) <223> HIV Protease

-75-

<2	21> 22> 23>	(298) ion	(105 of H	3) IV R	ever	se T	ranso	cript	tase						
CC	00> t ca o Gl	g at	c ac e Th	t ct r Le	t tgg u Trj	g caa p Glr	a cga n Arg	a cco	c cto Leu	ı Val	c aca	a ata r Ile	a arg	g ata a Ila 1!	a ggg e Gly	48
gg:	g caa y Gl	a cta n Lei	a aa u Ly 2	s Gli	a gct u Ala	t cta a Leu	tta Lei	a gat 1 Asp 25	Thr	a gga Gly	a gca ⁄ Ala	a gat a Asp	gat Asg 30	Th:	a gta r Val	96
tt: Le:	a gaa u Glu	a gaa 1 Glu 3!	u Mei	g aat t Asi	t ttg n Lev	g cca 1 Pro	gga Gly 40	/ Arg	tgg Trp	, aaa Lys	cca Pro	aaa Lys 45	Met	g ata : Ile	e Gly	144
Gl)	a att	e Glj	a ggt / Gl	t ttt y Phe	atπ Xaa	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	Gln	ata Ile	cyc Xaa	ata Ile	192
gaa Glu 65	ı Ile	tgt Cys	gga Gl	a yat 7 Xaa	aaa Lys 70	Ala	ata Ile	ggt Gly	acr Xaa	gta Val 75	Leu	gta Val	gga Gly	ccc Pro	acg Thr 80	240
ect	gto Val	aac Asn	rta Xaa	a att lle 85	Gly	aga Arg	aat Asn	ctg Leu	wtg Xaa 90	act Thr	cag Gln	att	ggt Gly	tgc Cys 95	Thr	288
tta Leu	aat Asn	ttt Phe	Pro 100	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	ttr Xaa	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aaa Lys	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtc Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	Gly 999	tta Leu	aag Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672

-76-

															kgg Xaa 240	720
att Ile	aga Arg	tay Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
								aaa Lys 265								816
caa Gln	aat Asn	cca Pro 275	gac Asp	cta Leu	gtt Val	att Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
								cat His								912
								ttt Phe								960
cag Gln	aaa Lys	gaa Glu	ccc Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	gca Ala	gtg Val 340	caa Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp		1053
<211 <212	> 43 > 10 > DN > Hu	82 A	Immu	modi	.fici	.ency	, Vir	rus (HIV)							
<222	> > CD > (0 > HI)	-	-												
<221 <222 <223	> (2	98).			' Rev	erse	Tra	nscr	ipta	se						
	caa							ccc Pro								48
Gly (96
tta (144

99 G1	a ati y Ile 50	e Gly	a ggt Y Gly	ttt Phe	ato E Ile	Lys 55	: Val	a aga L Arg	cag Gln	tat Tyr	gat Asp 60	Glr	g ata 1 Ile	ccc Pro	ata Ile	192
	u Xaa					: Ala					Lev				aca Thr 80	240
cc Pr	t gto o Val	aac L Asr	ata 1 Ile	att : Ile : 85	: Gly	aga Arg	aat Asn	ttg Leu	ttg Leu 90	Thr	cag Gln	att Ile	ggt	tgc Cys 95	Thr	288
tt: Le:	a aat u Asr	ttt Phe	e Pro	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
			Asp					aaa Lys					Thr			384
aaa Lys	a ata s Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aaa Lys	gaa Glu	Gly 999	432
aaa Lys 145	: Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc	ata Ile	aag Lys	aaa Lys	aag Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	ccg Pro	cat His 195	ccc Pro	gca Ala	Gly 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aag Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aaa Lys 225	Tyr	ast Xaa	gca Ala	ttt Phe	acc Thr 230	ata Ile	ccg Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	ccg Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	tgt Cys	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gaa Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864

-78-

gga tct gac ttg gaa ata ggg cag cat aga aca aaa ata gag gaa Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu 290 295 300	Leu
aga cag cat ctg ttg aaa tgg ggr ttt acc aca cca gac aag aaa Arg Gln His Leu Leu Lys Trp Xaa Phe Thr Thr Pro Asp Lys Lys 305 310 315	cat 960 His 320
cag aaa gaa cct cca ttc ctt tgg atg ggg tat gaa ctc cat cct Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro 325 330 335	gat 1008 Asp
aaa tgg aca gta caa ccg ata gag ctg cca gaa aaa gaa agc tgg Lys Trp Thr Val Gln Pro Ile Glu Leu Pro Glu Lys Glu Ser Trp 340 345 350	act 1056 Thr
gtc aat gac ata cag aag tta gtg gg Val Asn Asp Ile Gln Lys Leu Val 355 360	1082
<210> 44 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 44 cct cag atc act ctt tgg caa cga ccc atc gtc aca gta aag ata g Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Val Lys Ile G 1 5 10 15</pre>	gg 48 ly
ggg caa cta aag gaa gct yta tta gat aca gga gca gat gat aca g Gly Gln Leu Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Va 20 25 30	ta 96 al
tta gaa gaa atg aat tta cca gga aaa tgg aaa cca aaa ata ata gg Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Ile Ile Gl 35 40 45	99 144 ly
gga att gga ggt ttt gcc aaa gta aga cag tat gat cag ata ccc at Gly Ile Gly Gly Phe Ala Lys Val Arg Gln Tyr Asp Gln Ile Pro Il 50 55 60	ia 192 .e
	x 0
cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc ac Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Th 85 90 95	t 288 r

-79-

Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 1100 cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 aaa att tca aag att ggg cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 145 gcc ata aag aaa aaa aac agy act wga tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Asn Xaa Thr Xaa Trp Arg Lys Leu Val Asp Phe 175 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa ttr gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Xaa Gly 180 ata cca cat ccc tca ggg tta aaa aag aam aaa tca gta aca gta ctg Ile Pro His Pro Ser Gly Leu Lys Lys Xaa Lys Ser Val Thr Val Leu 195 gat gtg ggt gat gca tat ttt ca gtt ccc tta gat aaa gac ttc agg Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg 210 aaa tat act gca tta acc ata cct agt rta aac aat gag aca cca ggg Cty Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly 225 gca ata ttc caa agt agc at gt ct cca cag gga tgg aca agg aca cca ggg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly 225 gca ata ttc caa agt agc at gac at gac aca gga tcc cca ggg tra gac acca ggg tra gac at cca gac acca ggg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly 225 gca ata ttc caa agt agc at gac acca agg at cca cca ggg tac gac at ttc caa gat acca gac cca ggg tra gac at cca gac acca gac acca gac acca gac acca gac acca acca gac acca acca gac acca gac acca gac acca gac acca acca gac acca gac acca acca gac acca acca gac acca gac acca gac acca gac acca gac acca acca gac acca acca gac acca gac acca gac acca a																				
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga lys lis loo leu her fold glu lie cys lie lys Ala Leu val Glu lie cys Thr Glu Met Glu Lys Glu Gly liso liao liao liao liao liao liao liao lia		tt Le	a aa u As	nt t	he E	ro	att Ile	agt Sei	c CC	t at	e Gl	u Th	t gt r Va	a cc 1 Pr	a gt o Va	l Ly	s Le	a aag u Lys	Ī	336
aaa att tca aag att ggg cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 150 gcc ata aag aaa aaa aac agy act wga tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Lys Asn Xaa Thr Xaa Trp Arg Lys Leu Val Asp Phe 165 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa ttr gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Xaa Gly 180 ata cca cat ccc tca ggg tta aaa aag aam aaa tca gta aca gta ctg Ile Pro His Pro Ser Gly Leu Lys Lys Xaa Lys Ser Val Thr Val Leu 195 gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg aaa tat act gca ttt acc ata cct agt rta aac aat gag aca cca 210 gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg 210 aaa tat act gca ttt acc ata cct agt rta aac aat gag aca cca ggg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly 225 att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga cc cca Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro 245 gca ata ttc caa agt agc atg aca ag atc cta gag cct ttt aga aaa Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys 260 gga tct gat tta gaa ata gtg caa cat acg gg atg ga tgg ttg ctt gta Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val 275 gga tct gat tta gaa ata ggg caa cat aga gca aaa atc gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu 295 aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat Arg Gln His Leu Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca tcc tct tt gg atg ggt tt at gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp		CC Pr	a gg o Gl	y M	et A	sp (ggc Gly	Pro	a aaa b Lys	va.	L Ly:	a caa	a tg	g cca p Pro	o Le	u Th:	a ga r Gl	a gaa u Glu		384
gcc ata aag aaa aaa aac agy act wga tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Lys Asn Xaa Thr Xaa Trp Arg Lys Leu Val Asp Phe 165 170 170 177 178 179 189 190 175 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa ttr gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Xaa Gly 180 185 180 190 185 190 ata cca cat ccc tca ggg tta aaa aag aam aaa tca gta aca gta ctg Ile Pro His Pro Ser Gly Leu Lys Lys Xaa Lys Ser Val Thr Val Leu 20		aa Ly	s II	е г	aa g /s A	ca la	tta Leu	gta Val	. Glu	ı Ile	tgi Cy:	t aca	a gaa	u Met	Gl	a aaq u Lys	g gaa s Gla	a gga u Gly		432
Ala lie Lys Lys Lys Asn Xaa Thr Xaa Trp Arg Lys Leu Val Asp Phe 165 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt cat tr gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Xaa Gly 180 ata cca cat ccc tca ggg tta aaa aag aam aaa tca gta aca gta ctg 195		гĀ	3 II	t to e Se	ca a er L	ag a ys :	att Ile	Gly	Pro	gaa Glu	a aat a Asr	cca n Pro	Ty	r Asr	act	c cca	gta Val	l Phe		480
arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Xaa Gly 180 ata cca cat ccc tca ggg tta aaa aag aam aaa tca gta aca gta ctg 11e Pro His Pro Ser Gly Leu Lys Lys Xaa Lys Ser Val Thr Val Leu 200 gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gac ttc agg 67 Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg 210 aaa tat act gca ttt acc ata cct agt rta aac aat gag aca cca ggg 72 Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly 220 att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca 11e Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly 125 gca ata ttc caa agt agc atg aca aga atc cta gag cct ttt aga aaa Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys 260 cag aat cca gac ata gtt atc tat caa tac gtg gat gac ttg ctt gta Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val 270 gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg gga ctg gat gac ctg gat gac ctg Ser Pro 270 cag act cca gac ata gtt acc at caa tac gtg gat gac ttg ctt gta Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val 285 gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg ggg tct gat gac ctg Ile Glu Glu Leu 290 aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat gac Gln His Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp		gco	a Il	a aa e Ly	ıg a 's L	ys I	Lys	aac Asn	agy Xaa	act Thr	wga Xaa	Trp	Arg	a aaa g Lys	tta Lei	a gta 1 Val	. Asp	Phe		528
gat gtg ggt gat gca tat tt tca gtt ccc tta gat aaa gac ttc agg 67 Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg 210 aaa tat act gca ttt acc ata cct agt rta aac aat gag aca cca ggg 72 Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly 230 att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca 11e Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro 250 gca ata ttc caa agt agc atg aca aga atc cta gag cct ttt aga aaa Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys 265 gca ata tca gac ata gtt atc tat caa tac gtg gat gac ttg gat gac ttg gat gac aga aaca 61 Asp Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val 285 gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg gac ttg gat tta gaa ata 295 gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg gac gac ttg ctt gta 300 aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat Arg Gln His Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008 cag aaa gaa cct cca ttc ctt ttgg atg ggt tat gaa ctc cat cct gat 1008		aga Arg	gaa g Glu	a ct ı Le	u As	sn L	aag Lys	aga Arg	act Thr	caa Gln	Asp	Phe	tgg Trp	g gaa Glu	gtt Val	. Gln	Xaa	gga Gly		576
Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg 210 aaa tat act gca ttt acc ata cct agt rta aac aat gag aca cca ggg 72 Lys Tyr Thr Ala Phe Thr 11e Pro Ser Xaa Asn Asn Glu Thr Pro Gly 240 att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca 16e		ata Ile	cca Pro) HI	s Pi	c t	ca Ser	gly ggg	tta Leu	Lys	Lys	aam Xaa	aaa Lys	tca Ser	Val	Thr	gta Val	ctg Leu		624
Lys Tyr Thr Ala Phe Thr Ile Pro Ser Xaa Asn Asn Glu Thr Pro Gly 240 att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro 245 gca ata ttc caa agt agc atg aca aga atc cta gag cct ttt aga aaa Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys 260 cag aat cca gac ata gtt atc tat caa tac gtg gat gac ttg ctt gta Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val 275 gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu 290 aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat Arg Gln His Leu Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp		gat Asp	va.	. GI	t ga y As	it g ip A	jca la	tat Tyr	Phe	tca Ser	gtt Val	ccc Pro	tta Leu	Asp	aaa Lys	gac Asp	ttc Phe	agg Arg		672
gca ata ttc caa agt agc atg aca aga atc cta gag cct ttt aga aaa 816 Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys 260 cag aat cca gac ata gtt atc tat caa tac gtg gat gac ttg ctt gta 864 Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val 275 gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu 290 aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat Arg Gln His Leu Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp		гàг	туг	ac Th	t go r Al	a t a P	he	Thr	ata Ile	cct Pro	agt Ser	rta Xaa	Asn	Asn	gag Glu	aca Thr	cca Pro	Gly		720
Cag aat cca gac ata gtt atc tat caa tac gtg gat gac ttg ctt gta Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Leu Val 275 gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu 290 aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat Arg Gln His Leu Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp	:	att Ile	aga Arg	ta: Ty:	t ca r Gl	n T	yr .	aat Asn	gtg Val	ctt Leu	cca Pro	Gln	gga Gly	tgg Trp	aaa Lys	gga Gly	Ser	cca Pro		768
gga tct gat tta gaa ata ggg caa cat aga gca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Ala Lys Ile Glu Glu Leu 290 aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat Arg Gln His Leu Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp	į	gca Ala	ata Ile	Phe	GT.	n Se	gt a er a	agc Ser	atg Met	aca Thr	Arg	atc Ile	cta Leu	gag Glu	cct Pro	Phe	aga Arg	aaa Lys		816
aga caa cat ctg ttg agg tgg ggg ttt atc aca cca gac gaa aaa cat Arg Gln His Leu Leu Arg Trp Gly Phe Ile Thr Pro Asp Glu Lys His 305 cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp	Ċ	cag Gln	aat Asn	PIC	AS	c at p I]	re /	val	IIe	Tyr	caa Gln	tac Tyr	gtg Val	gat Asp	Asp	Leu	ctt Leu	gta Val		864
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 1008	Ġ	gga Sly	ser	gat Asp	tt: Le	a ga	aa a lu]	rie	Gly	caa Gln	cat His	aga Arg	gca Ala	Lys	ata Ile	gag Glu	gaa Glu	ctg Leu	;	912
Gin Lys Giu Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp		щg	caa Gln	cat His	Lei	g tt 1 Le	eu A	urg :	tgg Trp	ggg Gly	ttt Phe	atc Ile	Thr	cca Pro	gac Asp	gaa Glu	aaa Lys	His	9	960
	G	ag ln	aaa Lys	gaa Glu	Pro	PI	:0 F	tc o	ctt Leu	tgg Trp	atg Met	Gly	tat Tyr	gaa Glu	ctc Leu	cat His	Pro	gat Asp	10	800

-80-

aaa tgg aca gta cag ccc ata gtg ctg cca gaa aaa gay agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata caa aag tta gtg gga aaa ttg aat tgg gca agc cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat gca ggg Ile Tyr Ala Gly 370	1116
<210> 45 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 45 cct cag atc act ctt tgg caa cga ccc rtc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg cag cta aag gaa gct cta tta gat aca gga gca gac gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg aat tta cca gga aaa tgg aaa cca aaa atg ata gtg Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Val 35 40 45	144
gga att gga gga ttt gtc aaa gta aaa cag tat gag caa ata cct gta Gly Ile Gly Gly Phe Val Lys Val Lys Gln Tyr Glu Gln Ile Pro Val 50 55 60	192
gaa atc tgt gga cat aaa gct gta ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240
cct gcc aac ata att gga aga aat ctg ttg act cag att ggt tgc act Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca aaa gar Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Lys Glu 115 120 125	384

a: Ly	ys I.	ta m le %	naa (aa	gca Ala	tt Le	g gt u Va	a ga il Gl 13	u 11	it to le Cy	gt ad /s Tl	ca g nr G	lu	atg Met 140	: G1	a aa u Ly	g ga s G	aa lu	gga Gly	432
aa Ly ~14	5 1.	t t le S	ca er	aaa Lys	at:	t gg e Gl 15	y Pr	t ga o Gl	ia aa .u As	it co sn Pr	:0 T	ac yr 1	aat Asn	ac Th	t cc r Pr	a gt o Va	al :	ttt Phe 160	480
gc Al	t at .a I]	a a .e L	ag ys	aaa Lys	aaq Lys 165	s As	c ag n Se	t ga r As	t ag p Ar	a to g Tr 17	TO A	ga a rg 1	aaa Lys	tt: Le:	a gt ı Va	a.ga 1 As 17	p l	tc Phe	528
ag Ar	a ga g Gl	a c u L	eu	aat Asn 180	aag Lys	g ag s Ar	g ac g Th	t ca r Gl	a ga n As 18	p Ph	c to	.p (gaa Glu	att Ile	Car Gl: 190	n Le	a g	gga Sly	576
at Il	a cc e Pr	О п.	at is 95	ccc Pro	gca Ala	gg:	g tt. Y Le	a aa u Ly: 20:	a aa s Ly 0	g aa s Ly	g aa s Ly	a t	ca Ser	gta Val 205	Thi	a rt r Xa	a c a L	ta eu	624
ga Asj	t gt p Va 21	T G.	ly i	gat Asp	gca Ala	tat Ty:	Pho 215	e Se:	a rti r Xaa	t cc	c tt o Le	u A	at Asp 220	aaa Lys	gaa Glu	tt Ph	c a e A	gg rg	672
225	- Τ λ.	t ac	r Z	gca Ala	ttt Phe	acc Thr 230	. 116	e Pro	agt Ser	t ata	a aa = As 23	n A	at	gag Glu	aca Thr	CC:	o G	99 1y 40	720
att Ile	aga Arg	a ta g Ty	r (caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Glr 250	Gl	a t y T	rb aa	aaa Lys	gga Gly	Sei 255	r P	ca ro	768
Ald	TITE	: Pn	e G 2	660	Ala	Ser	Met	Thr	aaa Lys 265	Ile	. Le	ı G	lu	Pro	Phe 270	Arc	, L	ys.	816
GIII	ASI	27	5 5	ııu	Leu	vai	TTE	280	caa Gln	Tyr	' Val	L As	sp :	Asp 285	Leu	Tyr	Va	al	864
gga Gly	tct Ser 290	ga Asj	c t p L	ta (gaa Glu	ata Ile	gga Gly 295	cag Gln	cat His	aga Arg	aca Thi	a aa - Ly 30	ys :	ata Ile	gag Glu	gaa Glu	. ct	g u	912
305	GIU	nıs	5 L	eu i	Leu	110	Trp	GIY	tta Leu	Phe	Thr 315	Pr	co P	Asp	Gln	Lys	Hi 32	s 0	960
	D y S	GIL		3	325	PHE	Leu	Trp	atg Met	330 GTA	Tyr	Gl	u I	Leu	His	Pro 335	As	р	1008
275	1+1	****	34	10	3 T.11	PIO	Met	vai	ctg Leu 345	Pro	GIu	Ly	'S A	qa	Ser 350	Trp	Th	r	1056
gtc Val	aat Asn	gac Asp 355	Tie	a c eu G	ag a	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aa As:	n T	gg (1p /	gca Ala	agt Ser	Ca Gl	g n	1104

-82-

att tat cca ggg Ile Tyr Pro Gly 370	1116
<210> 46 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 46 cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aaa gta ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Val Gly 1 5 10 15</pre>	48
ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg aat ttg cca gga agg tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata tcc ata Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Ser Ile 50 55 60	192
gaa atc tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 85 90 95	288
tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
CCA gga atg gac ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta gta gag att tgt aca gaa atg gaa aag gaa gga Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135 140	432
aaa att tca aaa att ggg cct gaa aac cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 145 150 150	480

-83-

gc Al	c at a II	a a le I	ag ys	aaa Lys	aaa Lys 165	o Mo	c ag p Se	t ac r Th	t aa r Ly	ng tg 's Tr 17	P Ar	ga aa :g Ly	a tt s Le	a gt	a ga al As 17	at tto sp Phe	528
ag Ar	a ga g Gl	a c u L		aat Asn L80	aaa Lys	a aga	a ac	t ca r Gl:	a ga n As 18	p Pn	c tg e Tr	g ga p Gl	g gt u Va	t ca l Gl 19	n Le	a gga u Gly	576
at. Il	a cc e Pr	0 11	at d is B 95	ro	gca Ala	ggg Gly	g tta / Let	a aa 1 Ly: 200	s Ly	g aa s Ly	a aa s Ly	a tc s Se	a gt r Va 20	l Th	a gt r Va	a cta l Leu	624
gat Ās <u>ī</u>	t gt Va 21	- 0.	ly A	sp	gca Ala	tat Tyr	Phe 215	: Sei	a gti va:	t cco	c tta	a gat u Ası 220	o Gli	a ga u Asj	c tt p Ph	c aga e Arg	672
aaa Lys 225	2 -	t ac	t g r A	ca la	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	a aad e Asr 235	n Asr	gag 1 Glu	g aca	a cca Pro	a ggg Gly 240	720
act Thr	aga Arg	ta y Ty	t c T G		tac Tyr 245	aat Asn	gtg Val	Ctc Leu	cca Pro	cag Gln 250	r GTA	tgg Trp	, aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	tt Ph	~ O.	aa i ln (tgt Cys	agc Ser	atg Met	aca Thr	aaa Lys 265	TTE	tta Leu	gag Glu	cct	ttt Phe 270	Arc	aaa Lys	816
caa Gln	aat Asn	Pr 27		ac o	cta Leu	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	ga As _l	t t	a c u G	gaa Slu	ata Ile	gga Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ct Le	g t u L		agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	acc Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	CC Pr	-	ca (ro 1 25	ttt Phe :	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gt: Xaa 340	. G.	ag d ln E	ect a	ata Ile	val .	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc a		gac Asp 355	ata Ile	a ca e G	ag a ln L	ag t ys I	seu i	gtg g Val (gga Gly	aaa Lys	ttg Leu	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att (Ile 1	tac Tyr 370	cca Pro	ggg	•													1116

<210> 47 <211> 1116

-84-

	<212: <213:	DNA Hun	A man I	mmun	odif	icie	ncy	Viru	ıs (H	IV)						
	<220><221><222><223>	CDS (0)	; (297)												
<	:221> :222> :223>	(29	8)	.(11 of)	16) HIV :	Reve:	rse :	Fran	scrij	ptase	.					
С	400> ct c ro G	aa a	tc ac le Ti	et et hr Le	- 4 1 1	gg ca	aa co In Ar	ga co	CO TE	c gt eu Va	c ac	a gt ir Va	ca aa al Ly	/s I]	a gg e Gl	g 48 Y
g G	gg ca ly Gi	aa ci ln Le	,	ig ga rs Gl	a go .u Al	ct ct La Le	a tt u Le	u As	at ac sp Th	a gg ir Gl	a go y Al	a ga a As	p As	t ac p Th	a gta r Val	a 96 L
		3	5	c cy	5 <u>D</u> e	u FI	4	y AI	gır	ь гх	s Pr	o Ly 4	s Me 5	t Il	a ggg e Gly	,
G]	•	t gg e Gl 0	y Gl	t tt y Ph	t at e Il	c aa e Ly: 5:	o va	a ag l Ar	a ca g Gl:	a tai n Tyi	t ga r As _j	p G1:	g gt n Va	a gc l Al	c atg a Met	192
ga Gl 6	a at u Il 5	c tg e Cy	t gg: s Gl _:	a ca y Hi	t aag S Lys	2 VI	t ata a Ile	a gg e Gl	t aca y Thi	a gta r Val	L Le	a ata	a gg e Gl	a cci y Pro	aca Thr	240
cc Pr	t gte o Vai	c aa l Asi	c ata n Ile	a att = Ile 89	- 01	a aga / Arc	a aat J Asn	cto Lev	g ttg 1 Lev 90	Inr	caç Glr	g att	ggt Gly	tgo Cys	act Thr	288
			100		. Jei	. PIC	, ite	105	Thr	. val	Xaa	. Val	Lys 110	Leu	aag Lys	336
		115		 1		, arg	120	гуs	GIN	rrp	Pro	125	Thr	Glu	gaa Glu	384
aaa Lys	ata Ile 130	ara Xaa	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
	ata Ile	•	-1-	165		JC1	1111	пуѕ	170	Arg	ràs	Leu	Val	Asp 175	Phe	528
iga Irg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aaa Lys	act Thr	GIII	gac Asp 185	tty Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	ggr Xaa	576

-85-

ata Ile	ccg Pro	cat His 195	ccc Pro	gca Ala	ggg ggg	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctt Leu	624
gat Asp	gtg Val 210	gga Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gat Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aāā Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	att Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aag Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gga Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctr Xaa	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	ggg Gly	ytt Xaa	acc Thr	aca Thr 315	cca Pro	gac Asp	aag Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	ccy Xaa	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggk Xaa 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys '	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc a Val a	Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aar Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att i																1116
<210: <211: <212: <213:	> 11 > DN	A	Immu	nodi	fici	.ency	. Vir	us (HIV)							
<220: <221: <222: <223:	> CD > (0)	•	-												

-86-

<221> CDS <222> (298) ... (1115) <223> Portion of HIV Reverse Transcriptase cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg 48 Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 96 ata gaa gac ata gaa ttg cca gga aga tgg aaa cca aaa atg ata ggg Ile Glu Asp Ile Glu Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 144 gga att gga ggt ttt atc aaa gta aaa cag tat gag cag gta ccc ata 192 Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Glu Gln Val Pro Ile gaa ctc tgt ggg cgt aaa act ata ggt aca gta tta gta gga cct aca 240 Glu Leu Cys Gly Arg Lys Thr Ile Gly Thr Val Leu Val Gly Pro Thr cct gtc aac ata att gga aga aac ctg atg act cag att ggt tgc act 288 Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr 90 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag 336 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa 384 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa ggg 432 Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly aaa att tca aaa att ggg cct gaa aat cca tac aac act cca gta ttt 480 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe gcy ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc 528 Xaa Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe 170 aga gaa ctt aat aag aaa act caa gac ttc tgg gaa gtt caa tta gga 576 Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly ata cca cat cct gca ggg tta aaa aag aag aaa tca gta aca gta ttg Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu 624 gat gtg ggt gat gca tat ttt tca gtt ccg tta gat aaa gac ttc agg Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Asp Phe Arg 672

-87-

aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	TTE	cct Pro	agt Ser	ata Ile	aac Asn 235	. Asn	gag Glu	aca Thr	cca Pro	ggg Gly 240	720
~#. TTC	AIG	IÀT	GIN	245	. Asn	vai	Xaa	Pro	Gln 250	Gly	Trp	Lys	Gly	Ser 255		768
ΛIα	116	FIIE	260	ser	ser	Met	Thr	ьуs 265	Ile	Leu	gag Glu	Pro	Phe 270	Arg	Lys	816
caa Gln	ASII	275	Asp	Leu	vai	TTE	1yr 280	Gln	Tyr	Met	Asp	Asp 285	Leu	Tyr	Val	864
	290	Asp	ren	GIU	TTE	295	Gin	His	Arg	Thr	Lys 300	Ile	Glu	Glu	Leu	912
aga (Arg (305	3111	uis	Leu	Leu	310	Trp	GLY	Phe	Thr	Thr 315	Pro	Asp	Lys	Lys	His 320	960
cag a Gln I	ys ,	JIU	PLO	325	Pne	Leu	Trp	Met	330 GLY	Tyr	Glu	Leu	His	Pro 335	Asp	1008
aaa t Lys T	.rp.		340	GIN	PIQ	TTE	Val	Leu 345	Pro	Glu	Lys	Asp	Ser 350	Trp	Thr	1056
gtc a Val A	3	355	LIE (cag Gln	aag Lys	reu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	Asn '	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att t Ile S			ja													1115
<210><211><212><213>	111 DNA		mmur	nodi	ficie	ncy	Vir	18 (I	HIV)							
<220> <221> <222> <223>	(0)	(Pro	297) teas	e												
<221> <222> <223>	(29	B) Lion	.(11 of	16) HIV	Reve	rse	Tran	scri	ptas	ie						
<400> cct ca Pro Gl 1	gat	c ad le Tì	TT TI	tt t eu T 5	tb G aa c	aa c ln A	ga c rg P	ro L	tc g eu V 10	tc r al X	ca a aa I	ta a le L	ys I	ta g le G 15	jy gg	48

gg Gl	gg ca y Gl	g ci .n Le	eu Ly	ag ga /s Gl 20	a gc u Al	t ct a Le	a tt u Le	a ga u As 2	p Th	a gg r Gl	a gc y Al	a ga a As	t ga p As 3	p Th	a gta r Val	96
tt Le	a ga u Gl	u G.	aa at lu Me 15	g aa t As	t tte	g cc u Pr	a gg o Gl	a aga y Ara	a tgg g Tr	g aa o Lys	a cc s Pr	a aa o Ly 4	s Me	g at t Il	a ggg e Gly	144
gg G1	a at y Il 5	e G1	ja gg .y Gl	t tt y Ph	c ato e Ile	aaa Lys 5	s Val	a aga l Arg	a caq g Glr	g tat n Tyn	t gat r Asj 60	p Gl	g ata	a cc	c ata o Ile	192
ga Gl: 6	n TT	c tg e Cy	t gg s Gl	c ca y Hi:	t aaa s Lys 70	: Ala	t ata a Ile	a ggt e Gly	aca Thr	gta Val	l Lei	a gta ı Val	a gga L Gly	a cct / Pro	t aca Thr 80	240
Pro	t gte o Vai	c aa l As	c at n Il	a ati e Ile 85	e Gly	aga Arg	a aat g Asn	cta Leu	ttg Leu 90	Thr	cag Glr	g att n Ile	ggt Gl	tgo Cys	act Thr	288
tta Lei	a aat 1 Asi	tt n Ph	t cc e Pr 10	o Ile	agt Ser	cct	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aac Lys	Leu	aag Lys	336
Pro	a gga o Gly	a ate	t As	ggc Gly	cca Pro	aaa Lys	gtt Val 120	Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	: гл.	a gca s Ala	a tta a Leu	gta Val	gaa Glu 135	atc Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	Gly ggg	432
aaa Lys 145	TTE	tca Ser	a aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aa <u>c</u> Lys	aaa Lys	aag Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	·aag Lys	aaa Lys	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aam Xaa	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	acc Thr	gca Ala	ttt Phe	cca Pro 230	tcc Ser	cta Leu	gtt Val	Ile .	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	gga Gly 240	720
atc Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat (Asn	gtg Val	ctt Leu	Pro (cag Gln 250	gga Gly	tgg Trp	aaa Lys	Gly	tca Ser 255	cca Pro	768

-89-

gca a Ala I	ata [le	ttt Phe	caa Gln 260		agc Ser	atg Met	aca Thr	aaa Lys 265	tre	tta Leu	gag Glu	cct Pro	ttt Phe 270	Arg	aaa J Lys	816
caa a Gln A		cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	GIII	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga t Gly S 2:	er i	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	gta Val	gag Glu	gag Glu	ctg Leu	912
aga ca Arg G 305	•		Jeu	Deu	310	тър	GIY	Pne	Thr	315	Pro	Asp	Lys	Lys	His 320	960
cag aa Gln Ly				325		beu	TID	Met	330	lyr	Glu	Leu	His	Pro 335	Asp	1008
aaa tg Lys Tr	•		340				Val	345	PIO	GIU	гÀг	Asp	Ser 350	Trp	Thr	1056
gtc aa Val As	3.	55		cag Gln	aag Lys	LCU	gtg (Val (360	gga Gly	aaa Lys	tta Leu l	Asn :	tgg (Trp 2	gca Ala	agc Ser	cag Gln	1104
att tad Ile Ty: 370	r Pi	ca c	gly iaa													1116
<210> 5 <211> 1 <212> 1 <213> E <220> <221> C	1116 DNA Tuma	in I		odif	icie	ency	Viru	ıs (F	HIV)							
<222> (<223> H	IIV	(: Pro	297) teas	e												
<221> C <222> (<223> P	298) ion	(11: of 1	16) HIV 1	Reve:	rse :	Frans	scri	ptas	e						
<400> 50 cct cag Pro Gln 1	ato	-	5		.p G.	AI	.g Pi	:	ne va 10	al As	sn I	le L	ys I.	le G 15	ly	48
gga caa Gly Gln		2	0			.u <u>n</u> e	2	55 25	ır Gı	LY AI	la As	Sp As 3	p Th	ır Va	al	96
tta gaa Leu Glu	gaa Glu 35	at Me	g aa t As	t tt n Le	g cc u Pr	a gg o Gl 4	A wr	a to g Tr	g aa p Ly	a cc 's Pr	о гл	a tt s Le 5	g at u Il	a gg e Gl	ly 19	144

			e Gly					Val	aga . Arg				Gln				192
		Ile					Xaa		ggt Gly			Leu					240
						Gly			ctg Leu		Thr					Thr	288
					Ile				gaa Glu 105								336
				Asp					aaa Lys								384
			Lys						tgt Cys								432
									aat Asn								480
									aaa Lys								528
									gac Asp 185								576
									aag Lys								624
	gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	tat Tyr 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
3									agt Ser								720
	att Ile	aga Arg	tat Tyr	cag Gln	tay Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
									aga Arg 265								816
ć	aa Sln	Asn	cca Pro 275	gaa Glu	ata Ile	gtc Val	Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864

-91-

gca Ala	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gag Glu 295	Lys	cat His	aga Arg	a aca	a aa r Ly: 30	s Il	a ga e Gl	g ga u Gl	ia ci .u Le	tg eu	912
aga Arg ~~305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	tac Tyi	aca Thi	r Pro	a ga o As	c aa p Ly	a aa s Ly	g ca s Hi	İs	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	' Туз	gaa Glu	a ct 1 Le	c ca u Hi	t cc s Pr 33	o As	at sp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	Pro	gas Glu	aaa Lys	ga: Asj	e age p Sei 350	r Tr	g ac p Th	t	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aaa Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Tri	o Ala	a ag a Se:	t ca r Gl	g n	1104
att Ile	tat Tyr 370	gga Gly	ggg Gly														1116
<211 <212)> 51 > 11 > DN	A	Immu	nodi	fici	ency	· Vir	rus :	(HIV)							
<222	> CD: > (0) > HI)															
<222	> CDS > (29 > Por	98)	(1: 1 of	116) HIV	Rev	erse	Tra	nscr	ipta	ıse							
cct	> 51 cag a Gln 1	itc a [le]	et o	ctt i Leu :	tgg (Irp (caa o Gln <i>i</i>	ega Arg	ccc Pro	atc Ile 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	Gly	3	48
Gly (caa c Gln I	eu L	ag g ys 0 20	gaa g Slu <i>l</i>	gct o Na 1	cta t Leu I	ta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val		96
tta (gaa g Glu G	aa a lu M 35	tg a let A	at t sn I	tg d Leu F	ro G	ga a Sly i 40	aga Argʻ	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	ata Ile	ata Ile	ggg Gly		144
gga a Gly 1	att g [le G 50	ga g ly G	gt t ly P	tt a	ity a Kaa I	aa g ys V 55	yta a Yal A	aga (cag Sln	tat Tyr	gat Asp 60	cag Gln	ata Ile	cct Pro	ata Ile		192
gaa a Glu I 65	itc t	gt g ys G	ga c ly H	at a is L	aa a ys T 70	ct a hr I	ta g le G	ggt a Sly 1	aca (gta Val 75	tta Leu	ata Ile	gga Gly	cct Pro	aca Thr 80		240

Pr	O Al	a As	n II	e II 8	e Gly 5	y Ar	g As	p Le	u Le	u Th 0	r Gl	n Il	e Gl	у Сў 9	c act s Thr 5	
tt Le	a aa u As	t tt n Ph	t cc e Pr 10	O II	t agt e Sei	Pro	t ati	t gaa e Gli 10!	u Thi	t gt r Va	a cc l Pr	a gt. o Va	a aa l Ly: 110	s Le	g aag u Lys	336
Pro	a gg o Gl	a at y Me ll	t As	t gg p Gl	t cca y Pro	aaa Lys	a gtt s Val 120	l Lys	a caa s Glr	a tg	g cc p Pr	a tto D Lei 12!	ı Thi	a ga r Gl	a gaa u Glu	384
aaa Lys	a at: 5 Il: 13	е гу	a gca s Ala	a tta a Lei	a gta u Val	gaa Glu 135	ı Ile	tgt Cys	aca Thr	a gaa Glu	a cto 1 Let 140	ı Glı	a aag ı Lys	g gat s Asj	t ggg p Gly	432
aaa Lys 145	S TT6	t tca e Sei	a aaa r Lys	a att	ggg Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	: Ası	act Thr	cct Pro	gta Val	ttt l Phe 160	480
gcc Ala	ata Ile	a aag E Lys	g aaa E Lys	a aaa E Lys 165	: Asn	agt Ser	act Thr	aaa Lys	tgg Trp 170	Arc	aaa J Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	ı Lys	aga Arg	act Thr	caa Gln	gac Asp 185	Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	Pro	gcg Ala	Gly aaa	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	GTĀ	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	att Ile	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
gtt Val	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	ccc Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tat Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata (999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	Trp	agg (Arg : 310	rgg (Gly :	ttt Phe	Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960

-93-

ca Gl	g aa n Ly	a ga s Gl	a cco u Pro	C CC O Pro 325	o Phe	c ctt e Leu	tgg Tr	g ato Met	g ggt : Gl _y 330	, Tyr	gaa Glu	a cte	c car u Hi:	t cc s Pro	t gat o Asp 5	1008
aa Ly	a tg s Tr	g ac	a gta r Val 340	GL	a cct	ata Ile	gtg Val	g ctg Leu 345	. Pro	gaa Glu	aaa Lys	gad Asj	c ago o Sei 350	r Tr	g act p Thr	1056
gt Va	c aa l As:	t gad n Ası 35!	o Ile	caç Glr	g aaa 1 Lys	tta Leu	gtg Val 360	. Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	Ala	a agt	cag Gln	1104
		r Ala	a ggg a Gly													1116
<21 <21	10> 5 11> 1 12> I 13> F	L116 DNA	ı Imm	unod	ific	ienc	y Vi	rus	(HIV)						
<22	21> C 22> ((0)	.(29°	7) ase												
<22	1> C 2> (3> P	298)	(: on o	1116 E HI) V Re	verse	e Tra	ansci	ripta	ase						
cct	0> 5 caa Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	ctt Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	rta Xaa 15	gly ggg	48
ggg ggg	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta Leu	gaa Glu	Glu	atg Met	aat Asn	ttg Leu	cca Pro	gga Glv	aga	tgg	aaa	cca	aaa	atr	ata	āāa	144
		35					40	g	Trp	Lys	Pro	Lys 45	Xaa	IIe	GIÀ	
gga	att Ile 50	35 gga	ggt	ttt	atc	aaa Lys 55	40 qta	aga	caq	tat	gat	45 cag	ata	VCC	ata	192
gga Gly gaa	50 atc	gga Gly tgt	ggt Gly gga	ttt Phe cat	atc Ile aaa	aaa Lys	40 gta Val	aga Arg	cag Gln tca	tat Tyr	gat Asp 60 tta	45 cag Gln	ata Ile	ycc Xaa	ata Ile	192 240
gga Gly gaa Glu 65 cct Pro	of the state of th	gga Gly tgt Cys aac Asn	ggt Gly gga Gly ata	ttt Phe cat His aty Xaa 85	atc Ile aaa Lys 70 gga Gly	aaa Lys 55 gct	40 gta Val ata Ile aat	aga Arg ggt Gly ctg	cag Gln tca Ser atg Met	tat Tyr gta Val 75 act	gat Asp 60 tta Leu cag	cag Gln gta Val att	ata Ile gga Gly ggt Gly	ycc Xaa cct Pro tgc Cys 95	ata Ile aca Thr 80 act	

				Asp					. Lys					Thr		gaa Glu	384
_			Lys										Glu			Gly 999	432
		Ile							aat Asn								480
						Asn			aga Arg								528
									gac Asp 185								576
									aag Lys								624
									gtt Val								672
									agt Ser								720
									cca Pro								768
									aaa Lys 265								816
									caa Gln								864
	gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gaa Glu 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ttg Leu	912
									ttt Phe								960
									atg Met								1008
									ctg Leu 345								1056

-95-

gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 365	1104
att tay gca ggg Ile Tyr Ala Gly 370	1116
<210> 53 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 53 cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta aaa gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg aat tta cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gtg aga cag tat gat cag rta ccc ata Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Xaa Pro Ile 50 55 60	192
gaa att tgt gga cat aaa gct ata ggt aca gta tta gta gga tct aca Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Ser Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ctg ttg act cag ctt ggg tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta gta gaa att tgt aca gag atg gaa aag gaa ggg Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135 140	432

-96-

	: Ile										aat Asn					480
gco	ata a Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	Phe	528
aga Arg	a gaa g Glu	ctt Leu	aat Asn 180	Lys	aaa Lys	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
											tca Ser					624
		Gly									gat Asp 220					672
	Tyr										aat Asn					720
											tgg Trp					768
											gag Glu					816
aaa Lys	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
											aaa Lys 300					912
											cca Pro					960
											gaa Glu					1008
											aaa Lys					1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
		gca Ala														1116

-97-

<2 <2	12>	1116 DNA	n Im	nuno	dific	ciend	cy V:	irus	(HIV	7)						
<2.	21> 22>	CDS (0). HIV														
<2		(298)		(1116 of H)	5) IV Re	vers	se Tr	ransc	ript	ase						
cct	00> ! c cag c Glr	ato	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	aty Xaa 10	Val	aca Thr	ata Ile	aag Lys	g ata : Ile 15	ggg	48
G1 ⁷ 333	g caa ⁄ Glr	a cta n Leu	a aag Lys 20	Glu	gct Ala	yta Xaa	tta Leu	gat Asp 25	Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gta Val	96
tta Leu	gaa Glu	gac Asp 35) Met	gat Asp	ttg Leu	cca Pro	gga Gly 40	Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	Met	ata Ile	gtg Val	144
gga Gly	att Ile 50	Gly	ggt Gly	ttt Phe	gtc Val	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	ccc Pro	ata Ile	192
gaa Glu 65	Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	att Ile	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	ata Ile	gga Gly	aat Asn	aca Thr 80	240
cct Pro	gcc Ala	aac Asn	gta Val	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ctg Leu 140	gaa Glu	aag Lys	gat Asp	Gly aaa	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aag Lys 165	gac Asp	agt Ser	act Thr	Lys	tgg Trp 170	aga Arg	aaa Lys	gta Val	gta Val	gat Asp 175	ttc Phe	528

-98-

aga Arg	gaa Glu	ctt Leu	aac Asn 180	Lys	aga Arg	act Thr	caa Gln	gac Asp 185	Phe	tgg Trp	gag Glu	gtt Val	caa Gln 190	Leu	gga Gly	576
ata Ile	cca Pro	cac His 195	Pro	gca Ala	ggg	ata Ile	aaa Lys 200	Lys	aat Asn	aaa Lys	tca Ser	gta Val 205	act Thr	gta Val	cta Leu	624
gat Asp	gta Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	aga Arg	672
aaa Lys 225	Tyr	act Thr	gca Ala	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	att Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctc Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cac His	aga Arg	ata Ile	aaa Lys 300	ata Ile	rag Xaa	gaa Glu	ctg Leu	912
aga Arg 305	gaa Glu	cat His	cta Leu	tgg Trp	aag Lys 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aag Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	acg Thr	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	Gly 999	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile																1116

<210> 55 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)

<220>

-99-

<221> CDS <222> (0)...(297) <223> HIV Protease <221> CDS <222> (298)...(1116) <223> Portion of HIV Reverse Transcriptase <400> 55 48 cct caa atc act ctt tgg caa cga ccc ctc gtc gca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Ala Ile Lys Ile Gly ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gtc Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 96 tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 144 gga att gga ggt ttt atc aaa gta aag cag tat gat cag gta ctt gta Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Val Leu Val 192 gaa att tgt gga cat ara gct ata ggt aca gta tta gta gga cct aca 240 Glu Ile Cys Gly His Xaa Ala Ile Gly Thr Val Leu Val Gly Pro Thr 288 cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgt act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 336 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 105 cca ggt atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 384 aaa ata aaa gca tta ata gaa att tgt aca gaa atg gaa aag gaa ggg Lys Ile Lys Ala Leu Ile Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 432 130 aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt 480 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe gcc ata aag aaa aaa gac agt acc aaa tgg aga aaa tta gta gat ttc 528 Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe 170 aga gaa ctt aat aag aaa acg caa gac ttc tgg gaa gtt caa tta gga 576 Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly ata cca cat ccc gca ggg tta aaa aag aaa aaa tca gta aca gta ctg 624 Ile Pro His Pro Ala Gly Leu Lys Lys Lys Ser Val Thr Val Leu 200

-100-

gat gt Asp Va 21	lGly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aag Lys	gac Asp	ttc Phe	agg Arg	672
aag ta Lys Ty: 225	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att aga Ile Arq	a tat g Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctg Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca ata Ala Ile	ttt Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa aat Gln Ası	cca Pro 275	gac Asp	atg Met	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga to Gly Ser 290	: Asp	tta Leu	gaa Glu	ata Ile	gag Glu 295	cag Gln	cat His	aga Arg	rca Xaa	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
agg cag Arg Gli 305	g cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aag Lys	cat His 320	960
cag aaa Gln Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa tgg Lys Tr	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	ktg Xaa	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc aat Val Asr															1104
att tam Ile Xaa 370	Pro														1116
<210> 5 <211> 1 <212> E <213> E	116 NA	Immu	ınodi	.fici	.ency	, Vir	rus (HIV)							
<220> <221> C <222> (<223> H	0)														
<221> C <222> (<223> P	298).	(1 on of	116) HIV	'Rev	erse	: Tra	nscr	ipta	se						

-101-

	100>															
Pr 1	:0 G1	a at n Il	c ac	t ct r Le 5	t tg u Tr	g ca p Gl:	a cg n-Ar	a cco	c ato	e Va	c ac	a at r Il	a aa e Ly	g at s Il	a ggg e Gly 5	4.8
99 G1	g ca y Gl	a ct n Le	u Ly	g ga s Gl 0	a gc u Al	t cta a Lei	a tta u Lei	a gat u Asp 29	Th	a gga r Gly	a gca y Ala	a ga a As _j	t ga p As _l 30	o Th	a gta r Val	96
tt Le	a ga u Gli	u Gl	a at u Me 5	g aa t As:	t tte n Lei	g cca	a gga o Gly 40	/ Lys	tgg Tr	g aaa o Lys	a cca s Pro	a aaa b Lys 45	s Met	g ata	ggg Gly	144
gg Gl	a ati y Ile 50	e Gl	a gg y Gl	t tt: y Pho	t ato e Ile	aaa Lys 55	· Val	a aga L Arg	cag Glr	tat Tyr	gat Asp 60	Gli	g ata n Ile	a acc	ata Ile	192
ga Gl: 6.	u Ile	c tg e Cy	t gg s Gl	a cat y His	t aaa s Lys 70	. Ala	ata lle	ggt Gly	aca Thr	gta Val	Lev	gta Val	a gga L Gly	cct Pro	aca Thr 80	240
Pro	t gto o Val	aa Asi	c ata n Ile	a att = Ile 85	e Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	Thr	288
tta Le:	a aat u Asn	tt! Phe	Pro 100) Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336
Pro	a gga o Gly	ato Met	: Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	Lys	a gca s Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aaa Lys	gaa Glu	gly aaa	432
aaa Lys 145	att	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gat Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	Lys	aga Arg	act Thr	Gln	gac Asp 185	Phe	tgg Trp	gaa Glu	gta Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	61 y 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	cta Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720

-102-

att aga tat cag tac aat gtg ctt cca cag gga tgg aaa gga tca cca Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro 245 250 255	768
gca ata ttc caa agc agc atg aca aaa att tta gaa cct ttt aga aaa Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys 260 265 270	816
caa aat cca gaa ata gtt atc tat caa tac atg gat gat ttg tat gta Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val 275 280 285	864
gga tot gao tta raa ata gag cag cat aga aca aaa ata gag gaa otg Gly Ser Asp Leu Xaa Ile Glu Gln His Arg Thr Lys Ile Glu Glu Leu 290 295 300	912
aga caa cat ctg ttg agg tgg gga ttt acc aca cca gac aaa aag cat Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His 305 310 315 320	960
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata gtg ctg cca gaa cag gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Gln Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat cca ggg Ile Tyr Pro Gly 370	1116
<210> 57 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	D
<pre><400> 57 cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag tta ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Leu Gly 1 5 10 15</pre>	48
ggg caa cta atg gaa gtt cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Met Glu Val Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96

-103-

rta Xaa	gaa Glu	gaa Glu 35	ata Ile	agt Ser	tta Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly 999	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	gtc Val	aaa Lys 55	gta Val	aaa Lys	cag Gln	tat Tyr	gat Asp 60	cag Gln	gta Val	ccc Pro	tta Leu	192
gaa Glu 65	att Ile	tgt Cys	gga Gly	aaa Lys	aag Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gcc Ala	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ttt Phe	ttg Leu 90	gct Ala	cag Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttc Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gly ggg	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aag Lys 165	aac Asn	agt Ser	act Thr	aga Arg	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttt Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	agg Arg	acs Xaa	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	gly aaa	tta Leu	aar Lys 200	aag Lys	aac Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	Gly	Asp	Ala	Tyr	Phe	Ser	gtt Val	Pro	Leu	Asp	cca Pro	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cca Pro	ttt Phe 270	aga Arg	aaa Lys	816

-104-

caa aat cca Gln Asn Pro 275	gac ata Asp Ile	gtt at Val Il	c tgt e Cys 280	caa tad Gln Tyi	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga tct gac Gly Ser Asp 290	tta gaa Leu Glu	ata ga Ile Gl 29	u Gln 1	cat aga His Arg	Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga caa cat Arg Gln His 305	ctg ttg Leu Leu	agg tg Arg Tr 310	g gga t o Gly i	ttt tac Phe Tyr	aca Thr 315	cca Pro	gac Asp	caa Gln	aaa Lys	cat His 320	960
cag aaa gaa Gln Lys Glu	cct cca Pro Pro 325	ttc ct Phe Le	t tgg : 1 Trp	atg ggt Met Gly 330	Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa tgg aca Lys Trp Thr	gta cag Val Gln 340	cct at Pro Il	e Thr 1	ctg cca Leu Pro 345	gac Asp	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc aat gac Val Asn Asp 355	ata cag Ile Gln	aag tt Lys Le	a gtg g u Val (360	gga aaa Gly Lys	tta Leu	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att tat gca Ile Tyr Ala 370											1116
<210> 58 <211> 1116 <212> DNA <213> Human	Immunod	ificien	cy Vir	us (HIV	7)						
<220> <221> CDS <222> (0) <223> HIV Pr											
<221> CDS <222> (298). <223> Portio			se Tra	nscript	ase						
<400> 58 cct caa atc Pro Gln Ile	Thr Leu	Trp Gl:	n Arg 1	ccc cta Pro Lei 10	Val	aca Thr	ata Ile	aaa Lys	ata Ile 15	ggg ggg	48
ggg caa cta Gly Gln Leu	aag gaa Lys Glu 20	gct ct. Ala Le	a tta q ı Leu İ	gat aca Asp Thi 25	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta gaa gaa	atg act	ttg cc	a gga a	aaa tgg	aaa	cca	aaa	atg	ata	999 Gl v	144
Leu Ğlu Ğlu i 35	Met Thr	Leu Pr	40	ràs iri	, пув	FIO	45	1100	110	O17	

-105-

		Ile					Ala					. Leu				aca Thr 80	240
						Gly					Thr	cag Gln				Thr	288
					Ile					Thr		cca Pro					336
				Āsp					Lys			cca Pro		Thr			384
			Lys									atg Met 140					432
												aat Asn					480
	gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
												gaa Glu					576
												tca Ser					624
												gat Asp 220					672
1												aat Asn					720
1	att [le	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
Ē	gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	ata Ile 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
C	aa ln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
g	ly	tct Ser 290	gac Asp	tta Leu	gaa Glu	Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912

-106-

Arg Gln His Leu Leu Arg Trp Gly Phe Thr Thr Pro Asp Lys Lys His 305 310 315 320	960
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cca gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata aag ctg cca gac aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Asp Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat gca gga Ile Tyr Ala Gly 370	1116
<210> 59 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<400> 59	
cct caa atc act ctt tgg caa cga ccc tta gtc aca ata aag ata grg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Xaa 1 5 10 15	48
cct caa atc act ctt tgg caa cga ccc tta gtc aca ata aag ata grg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Xaa	48 96
cct caa atc act ctt tgg caa cga ccc tta gtc aca ata aag ata grg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Xaa 1 5 10 15 ggg caa cta aaa gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val	
cct caa atc act ctt tgg caa cga ccc tta gtc aca ata aag ata grg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Xaa 1	96
cct caa atc act ctt tgg caa cga ccc tta gtc aca ata aag ata grg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Xaa 1	96 144

-107-

t 1	ta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
	cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
ā	aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	rta Xaa	gaa Glu 135	atc Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	Gly 999	432
1	aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	aet Thr	cca Pro	gta Val	Phe 160	4.80
2	gcm Kaa	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
ä	aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtc Val	caa Gln 190	tta Leu	gga Gly	576
	ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	cta Leu	624
	gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gac Asp 220	caa Gln	gac Asp	ttc Phe	agg Arg	672
	aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
į	att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
	gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	agg Arg 265	atc Ile	tta Leu	gar Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
į	caa Gln	aat Asn	cca Pro 275	gaa Glu	ata Ile	gtc Val	aty Xaa	tat Tyr 280	cag Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	tta Leu	tat Tyr	gta Val	864
!	gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	gta Val	gag Glu	gaa Glu	ctg Leu	912
	aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agr Xaa 310	tgg Trp	Gly 399	ttt Phe	tmc Xaa	acg Thr 315	cca Pro	gac Asp	aaa Lys	aag Lys	cat His 320	960
,	cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	rap	1008

WO 01/35316 PCT/US00/30863

-108-

aaa tgg aca gta cag act ata gaa ctg cca gaa aaa gat agc tgg act Lys Trp Thr Val Gln Thr Ile Glu Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
ata tac cca ggg Ile Tyr Pro Gly 370	1116
<210> 60 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 60 cct caa atc act ctt tgg cag cga ccc cty gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta aaa gaa gct cta tta gay aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg aat ttg cca ggr aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Xaa Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata cct rta Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Xaa 50 55 60	192
gaa att tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ctg atg act cag ctt ggc tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Cys Thr 85 90 95	288
tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aga gtt aaa caa tgg cca ttg aca gaa gag Pro Gly Met Asp Gly Pro Arg Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384

aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	Glu	aag Lys	gaa Glu	gga Gly	432
	Ile					Pro					Asn				ttt Phe 160	480
					Āsp										ttc Phe	528
															gga Gly	576
		cat His 195														624
		ggt Gly													agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	acc Thr	aat Asn 235	aat Asn	gag Glu	aca Thr	ccm Xaa	999 Gly 240	720
		tat Tyr														768
		tty Phe														816
caa Gln	aat Asn	cca Pro 275	cac His	ata Ile	gtt Val	att Ile	ttt Phe 280	caa Gln	tat Tyr	gtg Val	gat Asp	gac Asp 285	ttg Leu	tat Tyr	gta Val	864
		gac Asp														912
		cat His														960
		gaa Glu														1008
		aca Thr														1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104

WO 01/35316 PCT/US00/30863

-110-

		Ala	a ggg a Gly													1116
 21> 21> س	0 > 6 1 > 1 2 > D 3 > H	116 NA	ı Imm	unoc	lific	ienc	y Vi	rus	(HIV	·)						
<22	1> C 2> (0)	.(29 rote	-												
<22		298)	on o) V Re	vers	e Tr	ansc	ript	ase						
cct		atc			tgg Trp					Val						48
				Ğlu	gct Ala											96
					ttg Leu											144
					atc Ile											192
					aag Lys 70											240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	mtt Xaa	ggt Gly	tgc Cys 95	act Thr	288
					agt Ser											336
					cca Pro										gaa Glu	. 384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	Gly 999	432
					999 Gly 150											480

-111-

	gco	a Il	a aa e Ly	g aa s Ly	a aag s Ly: 16!	s Asp	agt Ser	act Thr	aaa Lys	tgg Trp	o Arg	a aaa g Lys	a tta 5 Lei	a gta ı Val	a gat l Asr 175	ttc Phe	528	}
	aga Arg	ga:	a ct u Le	t aas u Ass 180	а Гуз	g aga s Arg	act Thr	caa Glr	gac Asp 185	Phe	tgg Trp	gaa Glu	a gtt 1 Val	cag Glr 190	Lev	a gga a Gly	576	i
	ata Ile	cca Pro	a car O His	s Pro	gca Ala	a ggg a Gly	tta Leu	aaa Lys 200	Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	. Thr	gta Val	ctg Leu	624	
	gat Asp	gto Val 210	r GT	gat Asr	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	Pro	tta Leu	gat Asp 220	Glu	agc Ser	ttc Phe	agg Arg	672	
	Lуs 225	Tyr	Thi	Ala	. Phe	acc Thr 230	Ile	Pro	Ser	Thr	Asn 235	Asn	Glu	Thr	Pro	Gly 240	720	
•	xaa	Arg	Tyr	GIn	245		Val	Leu	Pro	Gln 250	Gly	Trp	Lys	Gly	Ser 255	Pro	768	
4	чта	11e	Phe	260	Ser	agc Ser	Met	Thr	Lys 265	Ile	Leu	Glu	Pro	Phe 270	Arg	Lys	816	
(31n	Asn	275	Glu	Met	gtt Val	Ile	Tyr 280	Gln	Tyr	Met	Asp	Asp 285	Leu	Tyr	Val	864	
G	тУ	290	Asp	Leu	Glu	ata Ile	Glu 295	Gln	His	Arg	Thr	Lys 300	Ile	Glu	Glu	Leu	912	
3	05	GIN	HIS	Leu	Leu	agg Arg 310	Trp	Gly	Phe	Thr	Thr 315	Pro	Asp	Lys	Lys	His 320	960	
G	±π	ràs	GIu	Pro	9ro 325	ttt Phe	Leu	Trp	Met	Gly 330	Tyr	Glu	Leu	His	Pro 335	Asp	1008	
a L	aa ys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056	
y.	tc al	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu .	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104	
ai I	le :	tat Tyr 370	cca Pro	ggg ggg													1116	

<210> 62 <211> 1116 -112-

	12> I 13> I	-	ı Imn	unoc	lific	eienc	y Vi	rus	(HIV	')					
<22	21 > C 22 > C	(0)	.(29 Prote												
	2> ((298)	(.on c			vers	e Tr	ansc	ript	ase					
	0> 6			•											
														Gly 999	48
				Ğlu								gat Asp 30	Thr	gta Val	96
			Met									atg Met		Gly 999	144
		ĞĬy									Gln	ata Ile			192
	Ile											gga Gly			240
												ggc Gly			288
												aaa Lys 110			336
												aca Thr			384
												aag Lys			432
									Pro			cca Pro			480
												gta Val			528
												caa Gln 190			576

-113-

ata Ile	cca Pro	cat His 195	Pro	gca Ala	ggg	cta Leu	aaa Lys 200	aag Lys	aay Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtc Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
	tac Tyr															720
rtt Xaa	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
tca Ser	ata Ile	ttc Phe	caa Gln 260	tgt Cys	agc Ser	atg Met	acg Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
cag Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	trt Xaa 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gca Ala	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gag Glu 295	cag Gln	cat His	aga Arg	ata Ile	aaa Lys 300	ata Ile	gag Glu	gaa Glu	cta Leu	912
	caa Gln															960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gar Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Xaa	gtg Val 360	gga Gly	aaa Lys	ctg Leu	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
	tac Tyr 370															1116
<211 <212	> 63 > 11 > DN > Hu	16 A	Immu	nodi	fici	ency	Vir	us (HIV)							
<222	> > CD > (0 > HI)														

-114-

<221> CDS <222> (298) ... (1116) <223> Portion of HIV Reverse-Transcriptase cct caa atc act ctt tgg caa cga ccc gtt gtt aca gta agg ata ggg 48 Pro Gln Ile Thr Leu Trp Gln Arg Pro Val Val Thr Val Arg Ile Gly gga cag cta acg gaa gct yta tta gat aca gga gca gat gat aca gta 96 Gly Gln Leu Thr Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val tta gaa gaa atg act ttg cca gga aaa tgg aaa cca aaa ata ata ggg 144 Leu Glu Glu Met Thr Leu Pro Gly Lys Trp Lys Pro Lys Ile Ile Gly ggr att gga ggt ttt atc aaa gta aga cag tat gat cac gta ctt gta Xaa Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp His Val Leu Val 192 gaa atc tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca 240 Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr 75 ect gtc aac ata att gga aga aat ttg atg act cag ctt ggg ttc act 288 Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Leu Gly Phe Thr 85 90 tta aat ttt cca att agt cct att gaa act gta cca gta aaa tta aag 336 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 cca ggg atg gat ggc cca aaa gtt aaa caa tgg cca ttg mca gaa gaa 384 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Xaa Glu Glu aaa ata aaa gca cta aca gaa att tgt aca gaa ttg gaa aag gaa gga 432 Lys Ile Lys Ala Leu Thr Glu Ile Cys Thr Glu Leu Glu Lys Glu Gly aaa att tca aga ata ggg cct gaa aat cca tac aat act cca ata ttt 480 Lys Ile Ser Arg Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe gcc ata aag aag aaa aac ggt ayt agg tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Lys Asn Gly Xaa Arg Trp Arg Lys Leu Val Asp Phe 528 aga gag cta aat aag aga act caa gac ttc tgg gaa gtt caa cta gga 576 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 180 190 ata cca cat cct gca gga cta aaa aag aac aaa tca gta aca gta ctg 624 Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu gat gtg ggt gat gca tat ttt tca gtt ccc tta cat gaa gac ttt aga Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu His Glu Asp Phe Arg 672 210

WO 01/35316 PCT/US00/30863

-115-

	Tyr														gga Gly 240	720
					aat Asn										ccg Pro	768
					agc Ser											816
					gtt Val											864
					ata Ile											912
agg Arg 305	gaa Glu	cac His	cta Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	ttc Phe	acc Thr 315	cca Pro	gac Asp	gaa Glu	aag Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttt Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctt Leu	cat His	cct Pro 335	gat Asp	1008
					cct Pro											1056
					aag Lys											1104
	tat Tyr 370															1116
<211 <212)> 64 .> 11 !> DN !> Hu	16 A	Immu	nodi	fici	ency	Vir	us (HIV)							
<222	> CD > (0 > HI)														
<222	> CD: > (2 > Po:	98).	(1 n of	116) HIV	Rev	erse	Tra	nscr	ipta	se						
cct	> 64 cag a Gln i	atc a	act (Thr)	ctt Leu '	tgg (Irp (caa Gln	cga Arg	ccc Pro	atc Ile 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	Gly aaa	48

-116-

				Glu					Thr			gat Asp		Thr	gta Val	96
			Met					Lys				aaa Lys 45	Xaa		ggg	144
gga Gly	att Ile 50	Gly	99 <u>y</u> Xaa	ttt. Phe	rtc Xaa	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	syc Xaa	ata Ile	192
	Ile											gta Val				240
												ctt Leu				288
												gta Val				336
												tta Leu 125				384
												gaa Glu				432
												act Thr				480
gct Ala	ata Ile	aag Lys	aaa Lys	aag Lys 165	gac Asp	agt Ser	gct Ala	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
agg Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	Gly 999	576
ata Ile	Xaa	cat His 195	ccc Pro	gca Ala	ggg	Xaa	aaa Lys 200	Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	cta Leu	624
gat Asp	gta Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	caa Gln	aac Asn	ttc Phe	aga Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	ayg Xaa	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768

-117-

gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aa Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Ly 260 265 270	a 816 s
caa aat cca gar ata rtt atc tat caa tac gtg gat gat ttg tat gt Gln Asn Pro Glu Ile Xaa Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Va 275 280 285	a 864 1
gga tct gac ttr gaa ata ggg cag cat aga aca aaa ata gag gaa ct Gly Ser Asp Xaa Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Le 290 295 300	g 912 u
aga caa cat ytg ttg aag tgg gga ttt acc aca cca gac aag aag ca Arg Gln His Xaa Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His 305 310 315 320	3
cag aaa gaa cct cca ttc ctt tgg atg ggg tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata atg ctg cca gaa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Met Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Glr 355 360 365	j 1104
att tat gca gga Ile Tyr Ala Gly 370	1116
<210> 65 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 65 cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly 1</pre>	48
ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gac atc aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg Leu Glu Asp Ile Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly 35 40 45	144

-118-

		gga Gly														192
gac Asp 65	Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	Val	ata Ile	ggt Gly	aca Thr	gtg Val 75	Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct	gcc Ala	aac Asn	ata Ile	att Ile 85	ĞĪу	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
		ttt Phe														336
		atg Met 115														384
		aaa Lys														432
	Ile	tca Ser														480
		aag Lys														528 [.]
		ctt Leu														576
		cat His 195														624
		ggt Gly														672
Lys		act Thr	Āla	Phe		Ile	Pro	Ser	Ile	Asn	Asn					720
		tat Tyr														768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	att Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
		cca Pro 275														864

-119-

290 295 300	912
aga gaa cat ctg tgg aag tgg gga ttt tac aca cca gac aaa aaa cat Arg Glu His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His 305 310 315 320	960
cag aag gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata aag ytg cca gaa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Lys Xaa Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt caa Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat cca ggg Ile Tyr Pro Gly 370	1116
<210> 66 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297)	
<223> HIV Protease	
<223> HIV Protease <221> CDS <222> (298)(1116)	48
<223> HIV Protease <221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase <400> 66 cct cag atc act ctt tgg caa cga ccc ctt gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly	4 8 96
<pre><223> HIV Protease <221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase <400> 66 cct cag atc act ctt tgg caa cga ccc ctt gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1</pre>	
<pre><223> HIV Protease <221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase <400> 66 cct cag atc act ctt tgg caa cga ccc ctt gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1</pre>	96

-120-

cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	caa Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aga Arg	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gca Āla	gaa Glu 135	att īle	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	aat Asn	ara Xaa	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
agg Arg	gaa Glu	ctc Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	ggc Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aam Xaa	aaa Lys	tca Ser	gta Val 205	aca Thr	rta Xaa	ctr Xaa	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aar Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	wac Xaa 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	krc Xaa 245	aat Asn	gtg Val	yyt Xaa	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tcm Xaa 255	cca Pro	768
					agc Ser											816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tgt Cys 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ttg Leu	912
agg Arg 305	caa Gln	cat His	ttg Leu	ttg Leu	agg Arg 310	tgg Trp	ggr Xaa	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	ara Xaa	aaa Lys	cat His 320	960

WO 01/35316 PCT/US00/30863

-121-

cag aaa gag cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata aaa ctg cca gaa aaa gay agc tgg act Lys Trp Thr Val Gln Pro Ile Lys Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat gca ggg Ile Tyr Ala Gly 370	1116
<210> 67 <211> 1119 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1119) <223> Portion of HIV Reverse Transcriptase.	
<pre><400> 67 cct caa atc act ctt tgg caa cga cca ata gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
cta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt aty aaa gta aga cag tat gat cag ata tcc ata Gly Ile Gly Gly Phe Xaa Lys Val Arg Gln Tyr Asp Gln Ile Ser Ile 50 55 60	192
	240
gaa atc tgt ggg cat aaa gtt aca ggt aca gtg tta ata gga cct aca Glu Ile Cys Gly His Lys Val Thr Gly Thr Val Leu Ile Gly Pro Thr 65 70 75 80	
Glu Ile Cys Gly His Lys Val Thr Gly Thr Val Leu Ile Gly Pro Thr	288

cc Pr	a gg o Gl	a at y Me 11	t As	t ggo p Gly	c cca y Pro	a aaa o Lys	gtt Val	Lys	caa Glr	tgg Tr	g cca p Pro	tto Lev 129	ı Thi	a gaa c Glu	a gaa ı Glu	384
aa Ly	a ata s Ilo 130	e Ly:	a gca s Ala	a ttg a Lei	g gta 1 Val	gaa Glu 135	ι Ile	tgt Cys	gca Ala	gaa Glu	a ato 1 Met 140	: Gli	a aag 1 Lys	g gaa s Glu	a ggg a Gly	432
Gli 14	n Ile	tca Ser	a aaa c Lys	a att	gag Glu 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	aat Asr	cca Pro	gta Val	ttt Phe 160	480
gto Val	e ata l Ila	a aag E Lys	g aaa E Lys	aaa Lys 165	Asp	ggt	act Thr	aac Asn	tgg Trp 170	aga Arg	aaa Lys	tta Leu	ata Ile	gat Asp 175	ytc Xaa	528
aga Arg	a gaa g Glu	ctt Lev	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gat Asp 185	ttc Phe	tgg Trp	gaa Glu	att Ile	Caa Gln 190	Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aat Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	ttt Phe	tat Tyr 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gag Glu	aac Asn	ttc Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	atg Met 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	att Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
aac Asn	aat Asn	cca Pro 275	gac Asp	ata Ile	gtc Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gca Ala	tct Ser 290	gat Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	gaa Glu	cat His	cta Leu	ttr Xaa	aag Lys 310	tgg Trp	gga Gly	ttt Phe	Thr	aca Thr 315	cca Pro	gac Asp	aar Lys	aar Lys	yat Xaa 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ytc Xaa	ctt Leu	tgg Trp	Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata (Val	ctg (Leu 1 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056

-123-

) Ile					. Gly					Ala		cag Gln	1104
_		tat Tyr 370	Pro														1119
	<21 <21	0 > 6 1 > 1 2 > D 3 > H	119 NA	ı Imm	nunod	lific	ienc	y Vi	rus	(HIV)						
	<22	0> 1> C 2> (3> H	0)														
	<22	1> C 2> (: 3> P	298)				vers	e Tr	ansc	ript	ase						
	<40	0> 6	8														
	cct	caa	atc						ccc Pro								48
					Glu				gat Asp 25								96
									aaa Lys								144
									aga Arg								192
	gaa Glu 65	atc Ile	tgt Cys	Gly aaa	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gtr Xaa 75	tta Leu	ata Ile	gga Gly	ccc Pro	aca Thr 80	240
									ctg Leu								288
	tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
									aaa Lys								384
	aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	Ğlu	aag Lys	gaa Glu	gga Gly	432

-124-

aa: Ly: 14!	s Ile	tca Ser	a aaa Lys	att : Ile	999 Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gtt Val	ttt Phe 160	480
					Asp					Arg					ttc Phe	528
				Lys										Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	Pro	gca Ala	Gly	ttg Leu	aag Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	cta Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	aac Asn	ttt Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aat Asn 235	aat Asn	gaa Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
			caa Gln 260													816
caa Gln	aat Asn	cca Pro 275	gac Asp	atg Met	rtt Xaa	att Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
ggc	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gga Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gaa Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	Gly 999	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
			cct Pro			Leu	Trp		Gly	Tyr		Leu				1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gcg Ala	agt Ser	cag Gln	1104
			gly aaa													1119

-125-

<2 <2 <2 <2 <2: <2: <2:	20> 21> (22>	1119 DNA Huma: CDS (0).	n Imm (29 Prote		dific	ciend	cy V:	irus	/IH)	J)						
<2		(298)		(1119 of H		vers	e Tı	canso	ript	ase						
cct	00> 6 cag Glr	ato	c act	ctt Leu 5	tgg Trp	g caa Glm	cga Arg	ccc Pro	cty Xaa 10	Val	aca Thr	ata Ile	aag Lys	ata Ile	ggg	48
G1 ⁷	g caa g Glr	yta Xaa	aag Lys 20	Glu	gct Ala	mta Xaa	tta Leu	gay Asp 25	Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gtg Val	96
tta Lev	gaa Glu	gaa Glu 35	. Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	Ile	ata Ile	gjà aaa	144
gga Gly	att Ile 50	Gly	ggt	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	gag Glu	tat Tyr	gag Glu 60	cag Gln	ata Ile	caa Gln	gta Val	192
gaa Glu 65	Ile	tgt Cys	gga Gly	cat His	aag Lys 70	gct Ala	ata Ile	rgt Xaa	aca Thr	gta Val 75	tta Leu	ata Ile	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	cta Leu	atg Met 90	act Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gag Glu 105	act Thr	gta Val	ccg Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggt Gly	cca Pro	aga Arg	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	acy Xaa	ccr Xaa	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528

-126-

aga Arg	a gaa g Glu	ctt Leu	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Leu	gga Gly	576
ata Ile	ccg Pro	cat His 195	Pro	gca Ala	Gly	tta Leu	aag Lys 200	Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	ctr Xaa	624
gat	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tac Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Tle	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Prò	gga Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gaa Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	car Gln	tac Tyr	atg Met	gat Asp	gac Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	cta Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	tkg Xaa	agg Arg 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cac His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctr Xaa 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gcg Ala	agt Ser	cag Gln	1104
_	tat Tyr 370															1119

<210> 70 <211> 1119 <212> DNA

<213> Human Immunodificiency Virus (HIV)

<220>

WO 01/35316 PCT/US00/30863

-127-

<221> CDS <222> (0) ... (297) <223> HIV Protease <221> CDS <222> (298)...(1119) <223> Portion of HIV Reverse Transcriptase __<400> 70 48 cet caa ate act ett tgg caa ega eee ety gte kea ata aag gta ggr Pro Gln Ile Thr Leu Trp Gln Arg Pro Xaa Val Xaa Ile Lys Val Xaa ggg caa mta aag gaa gct yta tta gat aca gga gca gat gat aca gta Gly Gln Xaa Lys Glu Ala Xaa Leu Asp Thr Gly Ala Asp Asp Thr Val 96 tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 144 192 gga att gga ggt ttt atc aaa gta aaa cag tat gat cag gta arc ata Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Val Xaa Ile gaa atc tgt gga cat aaa gct ata ggt aca gta tta ata gga cct aca 240 Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr cct gtc aac ata att gga aga aay ctg ttg aca cag att ggt tgy act 288 Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag 336 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys cca gga atg gat ggc cca ara gty aaa caa tgg cca ttg aca gaa gaa 384 Pro Gly Met Asp Gly Pro Xaa Xaa Lys Gln Trp Pro Leu Thr Glu Glu 120 aaa ata aar gca tta atg gaa att tgt gca gay atg gaa aag gaa ggr Lys Ile Lys Ala Leu Met Glu Ile Cys Ala Asp Met Glu Lys Glu Xaa 432 aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 480 gcy ata aag aaa aaa gac agc act aaa tgg aga aaa tta gta gat ttc 528 Xaa Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe 165 aga gaa ctt aat aag aaa act caa gac ttt tgg gaa gtc caa tta gga 576 Arg Glu Leu Asn Lys Lys Thr Gln Asp Phe Trp Glu Val Gln Leu Gly ata cca cat ccy gca ggg tta aaa aag aac aaa tca gta aca gta ttg 624 Ile Pro His Xaa Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu 200

-128-

								gtt Val								672
								agt Ser								720
								cca Pro								768
								aaa Lys 265								816
								caa Gln								864
								cat His								912
								ttt Phe								960
cag Gln	aaa Lys	gaa Glu	ccc Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
								ttg Leu 345								1056
								gga Gly								1104
			Gly 999													1119
<211 <212	> 71 > 11 > DN > Hu	19 A	Immu	nodi	fici	ency	Vir	us (HIV)							
<222	> CD > (0)	(297 otea													
<222		98).	(1 n of			erse	Tra	nscr	ipta	se						

-129-

-40	0 - 7·	1											•			
cct	0> 7: caa Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	atc Ile 10	gtc Val	tca Ser	ata Ile	aag Lys	ata Ile 15	Gly 999	48
G1 y 999	gca Ala	aat Asn	aaa Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta Leu	gaa Glu	gaa Glu 35	atg Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aag Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	gtg Val	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	agc Ser	aaa Lys 55	gta Val	aga Arg	caa Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	ccc Pro	ata Ile	192
gaa Glu 65	atc Ile	tgc Cys	gga Gly	cgt Arg	aaa Lys 70	gtt Val	gta Val	ggt Gly	tca Ser	gta Val 75	tta Leu	ata Ile	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gcc Ala	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggc Gly	tgt Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	aaa Lys	gag Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	ata Ile	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	gaa Glu	gma Xaa	gga Gly	432
aaa Lys 145	att Ile	aca Thr	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	ccg Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	ata Ile	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aar Lys 165	aac Asn	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gac Asp 175	ttc Phe	528
agg Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	att Ile	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aar Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	acg Thr	aat Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720

-130-

					aat Asn										cca Pro	768
Āla					agc Ser											816
					gtt Val											864
					ata Ile											912
					aag Lys 310											960
					ttc Phe											1008
					cct Pro											1056
					aag Lys											1104
_	tat Tyr 370															1119
<211 <212 <213 <220 <221 <222		19 A man S)	(297)	fici.	ency	Vir	rus (HIV)							
<222		98).			Rev	erse	Tra	nscr	ipta	se						
<400 cct Pro	cag	atc Ile '	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	cty Xaa 10	gtc Val	aca Thr	ata Ile	aag Lys	atc Ile 15	Gly 999	48
ggg Gly																96

				ı Met					' Arg					Met		Gly ggg	144
	gga Gl _y	a ati	e Gly	a ggt / Gly	ttt Phe	rtc Xaa	aaa Lys 55	Val	aga Arg	caa Gln	tat Tyr	gat Asp 60	Gln	gta Val	ccc Pro	ata Ile	192
_	gaa Glu 65	ı Ile	tgo Cys	gga Gly	cat His	aaa Lys 70	Ala	ata Ile	ggt	aca Thr	gta Val 75	Leu	ata Ile	gga Gly	cct Pro	aca Thr 80	240
	ect Pro	gyo Xas	aac Asr	ata Ile	att Ile 85	gga Gly	aga Arg	aac Asn	ctg Leu	ttg Leu 90	act Thr	caa Gln	ctt Leu	ggc Gly	tgc Cys 95	act Thr	288
	tta Leu	aat Asn	ttt Phe	cca Pro 100	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
	cca Pro	gga Gly	atg Met 115	Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
	aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ctg Leu 140	gaa Glu	aaa Lys	gga Gly	agg Arg	432
	aaa Lys 145	aat Asn	tac Tyr	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aac Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
	gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
	aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
	ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	Gly 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
	gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aag Lys	gac Asp	ttc Phe	agg Arg	672
	aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agc Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
	att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctc Leu	Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
	gcm Xaa	ata Ile	ttc Phe	caa Gln 260	tgt Cys	agc Ser	atg Met	Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816

-132-

caa aat cca gaa ata gtt atc tat caa tac atg gat gat ttg tat gta Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val 275 280 285	864
ggg tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu 290 295 300	912
aga cga cat ctg ttg aag tgg gga ttt tac aca cca gac aaa aaa cat Arg Arg His Leu Leu Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His 305 310 315 320	960
cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gag ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta caa cct ata gtg cta cca gag aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aag tta aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
ata tac gca ggg att Ile Tyr Ala Gly Ile 370	1119
<210> 73 <211> 1119 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1119) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 73 cct caa atc act ctt tgg caa cga ccc ttc gtc aca gta aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Phe Val Thr Val Lys Ile Gly 1 5 10 15</pre>	48
ggg cag cta aag gaa gct cta tta gat aca gga gca gat aat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asn Thr Val 20 25 30	96
tta gaa gaa atg aat tta ccg gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga cag tat gat cag rta ccc ata Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Xaa Pro Ile 50 55 60	192

-133-

ga: Gl: 6!	u Ile	tgt Cys	gga Gly	a cad	c aaa s Lys 70	Ala	ata a Ile	a ggt e Gly	aca Thr	gta Val	Leu	a ata 1 Ile	gga Gly	cct Pro	aca Thr	240
Pro	gto Val	aac L Asn	ata Ile	att Elle 85	e Gly	aga Arg	ı gat J Asp	ctg Leu	ttg Leu 90	Thr	cac Glr	g ctt Leu	ggt Gly	tgo Cys	act Thr	288
tta Lei	a aat ı Asr	ttt Phe	Pro) Ile	agt Ser	Pro	att Ile	gat Asp 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Lev	aaa Lys	336
eca Pro	gga Gly	atg Met 115	Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	Lys	caa Gln	tgg Trp	cca Pro	Leu 125	Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	ctg Leu 140	gaa Glu	aag Lys	gaa Glu	GJA aaa	432
aag Lys 145	Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	acc Thr	cca Pro	gta Val	ttt Phe 160	480
gct Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aag Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gcg Ala	Gly 999	tta Leu	aaa Lys 200	aag Lys	aac Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	ccc Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	att Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
cag Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tac Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gac Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gat Asp	gag Glu	ctg Leu	912

-134-

						Trp					Pro				cat His 320	960
Gln					Phe	ctt Leu				Tyr					Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	Pro	gaa Glu	aar Lys	gac Asp	agc Ser 350	Trp	act Thr	1056
						tta Leu		Gly								1104
	tac Tyr 370															1119
<211 <212	0> 74 l> 11 2> DN 3> Hu	116 JA	Immi	ınod	ific:	iency	y Vi:	rus	(HIV))						
<222)> L> CI 2> (0 B> HI))														
<222	.> CI !> (2 !> Po	98).				verse	e Tra	ansci	ripta	ase						
cct	> 74 cag Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	ctc Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	gtc Val 15	gly aaa	48
Gly 999	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta (Leu (gag Glu	gaa Glu 35	cta Leu	aat Asn	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	gly aaa	144
ara :									C2.C	tat	ast	caq	ata	ccc	ata	192
Gly :	Ile 50	gga Gly	Gly	Phe	Ile	Lys 55	Val	Lys	Gln	Tyr	Asp 60	Gln	Ile	Pro	Ile	
gaa a Glu :	Ile (50 ata (Gly tgt	Gly gga	Phe cat	Ile aaa	Lys 55 gct	Val att	Lys ggt	Gln	Tyr gta	Asp 60 tta	Gln	Ile gga	Pro	Ile aca	240

-135-

					Ile					Thr					Leu	aag Lys	336
	ro								Lys			cca Pro		Thr			384
a	aa							Ile				atg Met 140					432
L												aat Asn					480
												aaa Lys					528
												gaa Glu					576
												tca Ser					624
												gat Asp 220					672
aa Ly 22	/S	tac Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	gga Gly 240	720
												tgg Trp					768
gc Al	a .a	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	agg Arg	aaa Lys	816
ca Gl	a .n	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tac Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
	y											aaa Lys 300					912
	g											cca Pro					960
ca Gl	g i	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gag Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008

-136-

aaa tgg ac Lys Trp Th				u Pro		Asp S			1056
gtc aat ga Val Asn As 35	p Ile Gl	g aag tta n Lys Lei	gtg gg Val Gl 360	a aaa y Lys	tta aat Leu Asn	tgg g Trp A 365	ca agt la Ser	cag Gln	1104
att tat go Ile Tyr Al 370									1116
<210> 75 <211> 819 <212> DNA <213> Huma	n Immunod	lificienc	y Virus	(HIV)					
<220> <221> CDS <222> (0). <223> Port		V Revers	e Trans	cripta	se				
<400> 75 ccc att ag Pro Ile Se 1									48
gat ggc cca Asp Gly Pro				Leu '		Glu Ly			96
gca ttg gta Ala Leu Val 3!	l Glu Ile	tgt aca Cys Thr	gaa atg Glu Met 40	g gaa : Glu :	aag gaa Lys Glu	gga aa Gly Ly 45	a att ⁄s Ile	tca Ser	144
aaa att ggg Lys Ile Gly 50	g cct gaa v Pro Glu	aat cca Asn Pro 55	tac aat Tyr Asr	act of	cca gta Pro Val 60	ttt go Phe Al	c ata a Ile	aag Lys	192
aaa aag gad Lys Lys Asp 65									240
aat aar aga Asn Lys Arg									288
ccc tca ggg Pro Ser Gly				Val 1			p Val		336
gat gca tat Asp Ala Tyr 115	Phe Ser								384
gca ttt acc Ala Phe Thr 130	ata cct Ile Pro	agt ata Ser Ile 135	aac aat Asn Asn	gag a Glu 1	aca cca Thr Pro 140	ggg at Gly Il	t agr e Xaa	tat Tyr	432

-137-

Cag Glr 145	Tyr	aat Asn	gtg Val	ctt Leu	cca Pro 150	Gln	gga Gly	tgg Trp	aaa Lys	gga Gly 155	Ser	cca Pro	gca Ala	ata Ile	ttc Phe 160	480
										Phe					cca Pro	528
gac	ata Ile	gtt Val	atc Ile 180	tat Tyr	caa Gln	tac Tyr	gtg Val	gat Asp 185	gat Asp	ttg Leu	tat Tyr	gta Val	gga Gly 190	Ser	gac Asp	576
tta Leu	gaa Glu	ata Ile 195	gag Glu	gag Glu	cat His	aga Arg	aca Thr 200	aaa Lys	ata Ile	gag Glu	gaa Glu	ctg Leu 205	agr Xaa	vrg Xaa	cat His	624
ctg Leu	tta Leu 210	aag Lys	tgg Trp	gga Gly	ttt Phe	acy Xaa 215	aca Thr	cca Pro	gac Asp	aaa Lys	aag Lys 220	cat His	cag Gln	aaa Lys	gaa Glu	672
cct Pro 225	cca Pro	ttt Phe	ctt Leu	tgg Trp	atg Met 230	ggt Gly	tat Tyr	gaa Glu	ctc Leu	cat His 235	cct Pro	gat Asp	aaa Lys	tgg Trp	aca Thr 240	720
gta Val	cag Gln	cct Pro	ata Ile	aag Lys 245	ctg Leu	cca Pro	gaa Glu	aaa Lys	gac Asp 250	agc Ser	tgg Trp	act Thr	gtc Val	aat Asn 255	gac Asp	768
ata Ile	cag Gln	aag Lys	tta Leu 260	gtg Val	gga Gly	aaa Lys	ttg Leu	aat Asn 265	tgg Trp	gca Ala	agt Ser	cag Gln	att Ile 270	tat Tyr	gca Ala	816
gly aaa																819
<211 <212 <213 <220 <221 <222)> 76 .> 81 !> DN !> Hu !> CD !> (0 !> Po	9 A man S	(819)												
CCC	> 76 att . Ile :	agt (Ser)	cct Pro	att Ile 5	gaa Glu	act Thr	gta Val	cca Pro	gta Val 10	aaa Lys	tta Leu	aag Lys	cca Pro	gga Gly 15	atg Met	48
gat Asp	ggc (Gly)	cca a Pro 1	aaa Lys : 20	gty Kaa	aaa Lys	caa Gln	tgg Trp	cca Pro 25	tta Leu	aca Thr	gaa Glu	gaa Glu	aaa Lys 30	ata Ile	aga Arg	96
gca Ala	tta q Leu V	gta g Val (35	gaa a Slu :	att Ile	tgt (Cys '	aca (gaa Glu i 40	atg (Met	gaa Glu	aag Lys	gaa Glu	gga Gly 45	aaa Lys	att Ile	tca Ser	144

WO 01/35316 PCT/US00/30863

-138-

aaa Lys	att Ile 50	Gly aaa	cct Pro	gaa Glu	aat Asn	cca Pro 55	tac Tyr	aat Asn	act Thr	cca Pro	gtg Val 60	ttt Phe	gct Ala	ata Ile	aag Lys	192	2
aaa Lys 65	aaa Lys	gac Asp	agt Ser	act Thr	aar Lys 70	tgg Trp	aga Arg	aaa Lys	ttg Leu	gta Val 75	gat Asp	ttc Phe	aga Arg	gaa Glu	ctt Leu 80	240)
 aat Asn	aag Lys	aga Arg	act Thr	caa Gln 85	gac Asp	ttc Phe	tgg Trp	gaa Glu	gtt Val 90	caa Gln	tta Leu	gga Gly	ata Ile	cca Pro 95	cat His	288	3
ccc Pro	tca Ser	gly ggg	tta Leu 100	aaa Lys	aag Lys	aaa Lys	aaa Lys	tca Ser 105	gta Val	aca Thr	gta Val	ctg Leu	gat Asp 110	gtg Val	ggt Gly	336	;
gat Asp	gca Ala	tat Tyr 115	ttt Phe	tca Ser	gtt Val	ccc Pro	tta Leu 120	gat Asp	aaa Lys	gac. Asp	ttc Phe	agg Arg 125	aag Lys	tat Tyr	act Thr	384	į
gca Ala	ttt Phe 130	act Thr	atn Xaa	cct Pro	agt Ser	ata Ile 135	aac Asn	aat Asn	gag Glu	aca Thr	cca Pro 140	Gly 999	att Ile	agg Arg	tat Tyr	432	:
cag Gln 145	tac Tyr	aat Asn	gtg Val	ctt Leu	cca Pro 150	caa Gln	gga Gly	tgg Trp	aaa Lys	gga Gly 155	tca Ser	cca Pro	gca Ala	ata Ile	ttc Phe 160	480)
caa Gln	agt Ser	agc Ser	atg Met	aca Thr 165	aaa Lys	atc Ile	tta Leu	gag Glu	cct Pro 170	ttt Phe	aga Arg	aaa Lys	caa Gln	aat Asn 175	cca Pro	528	ı
gac Asp	ata Ile	gtt Val	atc Ile 180	tat Tyr	caa Gln	tac Tyr	gtg Val	gat Asp 185	gat Asp	ttg Leu	tat Tyr	gta Val	gga Gly 190	tct Ser	gac Asp	576	;
cta Leu	gaa Glu	ata Ile 195	gga Gly	cag Gln	cat His	aga Arg	aca Thr 200	aaa Lys	ata Ile	gag Glu	gaa Glu	ctg Leu 205	aga Arg	cag Gln	cat His	624	•
ctg Leu	ttg Leu 210	agg Arg	tgg Trp	gga Gly	ttt Phe	acc Thr 215	aca Thr	cca Pro	gac Asp	aag Lys	aaa Lys 220	cat His	cag Gln	aaa Lys	gaa Glu	672	
cct Pro 225	ccc Pro	ttt Phe	ctt Leu	tgg Trp	atg Met 230	ggc Gly	tat Tyr	gaa Glu	ctc Leu	cat His 235	cct Pro	gat Asp	aaa Lys	tgg Trp	aca Thr 240	720	j
gta Val	cag Gln	cct Pro	ata Ile	gag Glu 245	ctg Leu	cca Pro	gac Asp	aag Lys	gat Asp 250	agc Ser	tgg Trp	act Thr	gtc Val	aat Asn 255	gac Asp	768	,
ata Ile	cag Gln	aag Lys	tta Leu 260	gtg Val	gga Gly	aaa Lys	tta Leu	aat Asn 265	tgg Trp	gca Ala	agt Ser	cag Gln	ata Ile 270	tat Tyr	gca Ala	816	
G1y 999																819	,

-139-

```
<210> 77
  <211> 1116
  <212> DNA
  <213> Human Immunodificiency Virus (HIV)
  <220>
__<221> CDS
  <222> (0)...(297)
  <223> HIV Protease
  <221> CDS
  <222> (298)...(1116)
  <223> Portion of HIV Reverse Transcriptase
  cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg
                                                                                      48
  Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly
                                                                                      96
  ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta
 Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val
 tta gaa gac atg aat ttg cca ggg aga tgg aaa cca aaa atg ata ggg
                                                                                     144
 Leu Glu Asp Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly
 gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata cct ata
                                                                                     192
 Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile
 gaa atc tgc gga cat aaa gct gta ggt aaa gta tta gta gga cct aca
Glu Ile Cys Gly His Lys Ala Val Gly Lys Val Leu Val Gly Pro Thr
                                                                                     240
 cct gtc aac ata att gga aga aat ctg ttg act caa ctt ggt tgc act
Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Leu Gly Cys Thr
                                                                                     288
 tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag
Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys
                                                                                     336
               100
 cca gga atg gat ggc cca aaa gtt aag caa tgg cca ttg aca gaa gaa
Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu
                                                                                    384
                                   120
 aaa ata aaa gca tta gta gaa att tgt aca gaa atg gag aag gaa ggg
                                                                                    432
 Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly
 aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt
                                                                                    480
 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe
                         150
                                                                                    528
 gct ata aag aaa aaa aac agt act aga tgg aga aaa tta gta gat ttc
 Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe
```

-140-

					Lys										Xaa	nnn Xaa		576
	nnn Xaa	nnn Xaa	nnn Xaa 195	nnn Xaa	nnn Xaa	gly ggg	twa Xaa	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu		624
	gat Asp	gta Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	gtt Val	cct Pro	cta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	aga Arg		672
	aag Lys 225	tac Tyr	act Thr	gca Ala	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240		720
	att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctg Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro		768
	gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys		816
	caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gtg Val		864
	gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	:	912
	aga Arg 305	caa Gln	cat His	ttg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	9	960
	cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	10	800
	aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	10	56
9	gtc Val	Asn .	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	Asn	tgg Trp 365	gca Ala	agc Ser	cag Gln	11	L04
	Ile :	tat Tyr 370	gca Ala	gly aaa													11	116
								_										

<210> 78 <211> 1122 <212> DNA <213> Human Immunodificiency Virus (HIV)

-141-

<2	22>	(0).	(2 Prot	297) :ease	:												
<2	21> 22> 23>	(298) ion	(112 of H	2) IV R	ever	se T	rans	crip	tase						·	
CC	00> t ca o Gl	gat	c ac e Th	t ct r Le 5	t tg u Tr	g caa p Gli	a cg n Ar	a cc g Pr	c cto o Leo 10	ı Va	c ac l Th	a at r Il	a aa e Ly	g at s Il l	a ggg e Gly 5	48	
999 Gl	g ca y Gl	a ct n Le	a aa u Ly 2	S GI	a gc u Ala	t cta a Lev	a tta 1 Lei	a gat u Ası 2!	p Thi	a gga c Gly	a gca y Ala	a ga a As	t ga p As _l 30	p Th:	a gta r Val	96	
tt: Le:	a gaa 1 Gl	a gad 1 Asj 3!	b we.	g gat t As _l	t ttg D Lei	g cca	a gga o Gly 40	/ Arg	a tgg g Trp	g aaa b Lys	a cca s Pro	a aaa o Lys	s Met	g ata	a ggg e Gly	144	
gga Gl	a att	: GTZ	a ggt y Gly	t tti Y Phe	ato E Ile	aaa Lys 55	Val	a aaa Lys	a cag s Glr	tat Tyr	gat Asp	Glr	g ata n Ile	a cco	ata Ile	192	
gaa Glu 65	1 116	tgt Cys	gga Gly	a cat / His	aaa Lys 70	Val	ata Ile	ggt Gly	aca Thr	gta Val 75	Lev	gta Val	gga Gly	cct Pro	aca Thr 80	240	
cct Pro	gto Val	aac Asn	ata Ile	att Elle 85	GTA	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288	
tta Leu	aat Asn	ttt Phe	Pro 100	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336	
cca Pro	gga Gly	atg Met 115	gat Asp	ggc	cca Pro	aga Arg	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384	
aaa Lys	ata Ile 130	aaa Lys	gca Ala	ttg Leu	gta Val	gaa Glu 135	ata Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gly aaa	432	
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	acr Xaa	cca Pro	∙gta Val	ttt Phe 160	480	
gcc Ala	ata Ile	arg Xaa	aaa Lys	aaa Lys 165	gaa Glu	agc Ser	tct Ser	agc Ser	tct Ser 170	aaa Lys	tgg Trp	aga Arg	aaa Lys	tta Leu 175	gta Val	528	
gat Asp	ttc Phe	aga Arg	gaa Glu 180	ctt Leu	aat Asn	aar Lys	aga Arg	act Thr 185	caa Gln	gac Asp	ttt Phe	ttk Xaa	gaa Glu 190	gtt Val	caa Gln	576	
tta Leu	GIA	ata Ile 195	cca Pro	cat His	ccc Pro	gca Ala	999 Gly 200	tta Leu	aag Lys	aag Lys	Lys	aaa Lys	tca Ser	gya Xaa	aca Thr	624	

-142-

rta Xaa	ttg Leu 210	. Asj	t gt p Va	9 99 1 Gl	t gat y As <u>r</u>	gc Ala 21	а чу	t tti	t tc: e Se:	a gti r Val	cco l Pro 220	o Let	a gat 1 Asp	raa Xaa	a gac a Asp	672
ttc Phe 225	agg Arg	aag Lys	g tat s Typ	act Th:	gca Ala 230	r Pile	t acc	ata Ile	a cct	agt Ser 235	: Ile	a aac e Asn	aat Asn	gag Glu	aca Thr 240	720
cca Pro	Gly	att Ile	aga Arg	tat Tyr 245	. G111	tac Tyr	aat Asn	gto Val	t ctt Leu 250	Pro	cag Gln	gga Gly	tgg Trp	aaa Lys 255	gga Gly	768
tca Ser	cca Pro	gct Ala	ata Ile 260	FILE	caa Gln	agt Ser	agc Ser	atg Met 265	Thr	aaa Lys	atc Ile	tta Leu	gag Glu 270	cct Pro	ttt Phe	816
aga Arg	aaa Lys	caa Gln 275	aat Asn	cca Pro	gay Asp	ata Ile	gtt Val 280	atc Ile	tat Tyr	caa Gln	tac Tyr	atg Met 285	gat Asp	gat Asp	ttg Leu	864
tat q Tyr i	gta Val 290	gga Gly	tct Ser	gay Asp	tta Leu	gaa Glu 295	ata Ile	gag Glu	cag Gln	cat His	aga Arg 300	ata Ile	aaa Lys	ata Ile	gag Glu	912
gaa d Glu I 305	etg Leu	aga Arg	caa Gln	yat Xaa	ytg Xaa 310	tgg Trp	arg Xaa	tgg Trp	ggr Xaa	ttt Phe 315	tac Tyr	aca Thr	cca Pro	gac Asp	aaa Lys 320	960
aaa c Lys H	at Iis (cag Gln	aaa Lys	gaa Glu 325	cct Pro	cca Pro	ttc Phe	cat His	tgg Trp 330	atg Met	ggt Gly	tat Tyr	Glu	ctc Leu 335	cat His	1008
cct g Pro A	at a sp 1	-3 -	tgg Trp 340	aca Thr	gta Val	cag Gln	PIO	ata Ile 345	gtg Val	ctg Leu	cca Pro	Glu	aaa Lys 350	gac Asp	agc Ser	1056
tgg a Trp T		tc /al /55	aat Asn .	gac Asp	ata (Ile (3111	aag Lys : 360	tta Leu	gtg Val	gga Gly	Lys :	ttg / Leu / 365	aat Asn '	tgg : Irp :	gca Ala	1104
agt ca Ser G 3	ag a ln I 70	itt (tat (Tyr)	gca Ala	ggr Kaa											1122
<210><211><212><213>	111 DNA		mmur	odif	icie	ncy	Viru	ıs (F	IIV)							
<220><221><222><223>	CDS	(297)													
<221> <222> <223>	(298	3) ion	.(11 of	16) HIV	Reve:	rse	Tran	scri	ptas	e						

-143-

	400>																		
	ct c ro G 1	ag ln	ato	act Thi	t ct r Le 5	t tg u Tr	g ca p Gl	a cg n Ar	a cc g Pr	c ct o Le l	c gt u Va 0	t ac l Th	a at r Il	a aa e Ly	s V	ta al 15	gly aaa	4	8
g G	gg ca ly G	aa ln	cta Leu	aag Lys 20	S GI	a gc u Al	t ct a Le	a tt u Le	a ga u As _j 2!	p Th	a gga r Gl	a gc y Al	a ga a As	c aa p As 3	n Tl	ca	gta Val	9	6
t i	ic ga ne Gl	aa d Lu	gac Asp 35	тел	ga: Asj	t tta p Le	a cca u Pro	a gg o Gl:	y Arg	g tgg g Trj	g aaa o Lys	a cc	a aa o Ly 4	s Me	g at t I]	a le	gly ggg	14	4
gg GI	.y 11	t d e d	gga Gly	ggt Gly	ttt Phe	t ato	c aaa e Lys 55	s va.	a aaa l Lys	a caq s Glr	g tat	gag Glu	a Gli	g ata n Ile	a co	:c	ata Ile	192	2
Gı	a at u Il 5	c t e C	gt Cys	gly	cgt Arg	aaa J Lys 70	A A La	ata Ile	a ggt e Gly	aca Thr	gtg Val	Let	a gta ı Val	a gga L Gly	a co / Pr	t i	aca Thr 80	240	כ
Pr	t gt o Va	c a l A	ac Asn	ata Ile	att Ile 85	GIY	aga Arg	gat Asp	ctg Leu	ttg Leu 90	act Thr	Cag Glm	g att 1 Ile	ggt Gly	, Су	c a s :	act Thr	288	}
ct Le	a aa u As:	t t	tt	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Le	a a u I	aag	336	i
Pre	a gga	Λ IAI	tg et 15	gat Asp	ggc Gly	cca Pro	aga Arg	gtt Val 120	Lys	caa Gln	tgg Trp	cca Pro	Leu 125	Thr	ga: Gl:	a g u G	jaa Slu	384	
aaa Lys	a ata 5 Ile 130	= 10	aa ys	gca Ala	tta Leu	ata Ile	gaa Glu 135	att Ile	tgt Cys	gca Ala	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaz Glı	a 9	lgg 1	432	
aaa Lys 145	, 116	to Se	ca er	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	L P	tt he 60	480	
gcc	ata Ile	aa Ly	ag a ys 1	uys	aag Lys 165	aac Asn	agt Ser	aat Asn	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	P	tc he	528	
aga Arg	gaa Glu	Le	u z	aat Asn 180	aag Lys	aga Arg	Thr	GIn	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	g G	ga ly	576	
ata Ile	cca Pro	ca Hi 19	. D .	ecc (gca Ala	Gly aaa	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aag Lys	tca Ser	ata Ile 205	aca Thr	gta Val	t i	ta eu	624	
gat Asp	gtg Val 210	gg Gl	rt g y A	at g	gca Ala	ıyr	ttc Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	ag Az	:g	672	
aag Lys 225	tat Tyr	ac Th	t g r A	ca t la E	-116	acc Thr 230	ata Ile	cct Pro	agt Ser '	Thr .	aac a Asn 1 235	aat Asn	gag Glu	aca Thr	cca Pro	gg G1 24	Y	720	

-144-

att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	Asn	gtg Val	ctg Leu	cca Pro	cag Gln 250	Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	Ile	tta Leu	gag Glu	cct	ttt Phe 270	aga Arg	aaa Lys	816
Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gat Asp	tta Leu	gaa Glu	ata Ile	gag Glu 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gat Asp	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ctt Leu	acc Thr	aca Thr 315	cca Pro	gac Asp	cag Gln	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gac Asp	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	ggr Xaa	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	caa Gln	1104
	tac Tyr 370															1116
<211 <212)> 80 .> 11 !> DN !> Hu	19 A	Immu	nodi	fici	ency	Vir	us (HIV)							
<222	> CD > (0 > HI)														
<222	> CD: > (2 > Po:	98).			Rev	erse	Tra	nscr	ipta	se						
cct	> 80 cag a Gln :	atc a Ile :	act o	ctt Leu 5	tgg (Trp (caa (Gln)	cga Arg	ccc Pro	ctc Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	gly ggg	48
gly aaa	cag (Gln 1	cta a Leu I	aag g Lys (20	gag Slu	gct (Ala 1	cta 1 Leu 1	tta (Leu i	gat a Asp 1 25	aca (Thr	gga (Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96

-145-

Le	a ga u Gl	a ga u Gl 3	u Me	g aat t Asi	t tte	g cca u Pro	gga Gly	Arc	tgg Tr	g aaa D Lys	a cca s Pro	a aaa b Lys 45	s Met	g ata	ggg Gly	144
Gl	a att	≥ Gl	a ggt y Gl	t ttt y Phe	t ato	c aaa e Lys 55	Va]	a aga L Arg	cag Glr	g tat 1 Tyr	gat Asp 60	Glr	g ata n Ile	a cto	ata Ile	192
Gl	a att u Ile 5	tg: Cy:	t gga s Gly	a cat / His	aaa Lys	s Ala	ata Ile	ggt Gly	aca Thr	gta Val	. Leu	gta Val	a gga . Gly	a cct	aca Thr 80	240
cc Pr	t gto o Val	aad Asi	c ata n Ile	a att	Gly	a aga / Arg	aat Asn	ctg Leu	ttg Leu 90	Xaa	cag Gln	att Ile	ggt Gly	tgc Cys 95	Thr	288
tt Le	a aat u Asr	ttt Phe	ccc Pro 100	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336
Pr	a gga o Gly	Met	Asp	ggc Gly	cca Pro	aga Arg	gtt Val 120	Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	Thr	gaa Glu	gaa Glu	384
aaa Ly:	a ata s Ile 130	Lys	a gca s Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aar Lys	gaa Glu	gj aaa	432
aaa Lys 145	a att s Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	rta Xaa	ttt Phe 160	480
gco	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arc	gaa Glu	ctt Leu	aat Asn 180	aag Lys	agg Arg	act Thr	caa Gln	gat Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	Gly 999	ttg Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttc Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	aga Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816

-146-

caa aat cca gac ata gtc atc tat caa tac atg gat gat ttg tat gta Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val 275 280 285	864
gga tct gac tta gaa ata ggg cag cat agg aca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu 290 295 300	912
aga caa cat ttg ttg aag tgg ggg ttt acc aca cca gac aaa aaa cat Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His 305 310 315 320	960
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gtg cag cct ata gtg tta ccg gaa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tac cca ggg att Ile Tyr Pro Gly Ile 370	1119
<210> 81 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 81 cct caa atc act ctt tgg caa cga ccy ctt gtt rcc ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Xaa Leu Val Xaa Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta arg gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Xaa Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa ata aat ttg cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aaa cag tat gat caa ata ccy rta Gly Ile Gly Gly Phe Ile Lys Val Lys Gln Tyr Asp Gln Ile Xaa Xaa 50 55 60	192

-147-

G	aa a lu I 65	tt t le C	gt Ys	gga Gly	ca: Hi	s Ar	ga go gg Al	et at la II	a gg Le G	gt a Ly :	aca Thr	gtv Xaa 75	a Le	a gt u Va	a gg	ga d ly I	Pro	aca Thr 80	•	240
Co Pi	ct gt ro Va	al A	ac sn	ata Ile	att Ile 89	e GI	a ag y Xa	r aa a As	it ct in Le	g t	teu Seu	act Thr	Gl:	g at n Il	t gg e Gl	yt t	gc ys 95	act Thr	•	288
tt L∈	a aa eu As	at t sn P	ne	ccc Pro 100	att Ile	ag Se	t co r Pr	t at o Il	t ga e Gl 10	u 1	ct	gta Val	cca Pro	a gt o Va	a aa l Ly 11	s L	ta eu	aag Lys		336
Pr	a gg o Gl	A M	tg et 15	gat Asp	ggc	Pro	a aa o Ly	a gt s Va 12	l Ly	a c s G	aa In	tgg Trp	Pro	tte Le 12	u Th	a g r G	aa lu	gaa Glu		384
aa Ly	a at s Il 13	е п	aa y	gca Ala	ttg Leu	gta Va	a ga l Gli 13!	n II	t tg e Cy	t a s T	ca	gaa Glu	ato Met	: Glı	a aa u Ly	g g s G	aa lu	gga Gly		432
аа L y 14	a at s Il	t to e Se	ca a	aga Arg	att Ile	999 Gly 150	Pro	ga: Gl:	a aa u As:	t c	ro	tac Tyr 155	aat Asn	act Thi	r Pro	ag:	ta al	ttt Phe 160		480
gc Al	t ata	a aa e Ly	rs I	aaa Lys	aar Lys 165	gat Asp	agt Sei	act Thi	t aaa Lys	5 T:	99 70	aga Arg	aaa Lys	tta Lei	a gta ı Val	l As	at sp 75	ttc Phe		528
ag Ar	g gaa g Glu	a ct ı Le	u P	aat Asn 180	aag Lys	agg Arg	act Thr	caa Glr	a gad Asp 185	P	tc ne	tgg Trp	gaa Glu	gtt Val	caa Glr 190	ı Le	a eu	gga Gly		576
110	e Pro	19	5 5	TO.	Ата	GIĀ	Leu	200	Lys	: L}	/S]	Lys	Ser	Val 205	Thr	Va	1	Leu		624
vəř	yte Val 210	. GI	уд	sp A	Ala	ıyr	215	Ser	Val	. Pr	o I	Leu	Asp 220	Lys	Glu	Ph	e i	Arg		672
aag Lys 225	tat Tyr	act	t g r A	ca t la 1	Pne	act Thr 230	ata Ile	cct Pro	agt Ser	ac Th	r A	aac Asn 235	aat Asn	gag Glu	aca Thr	Pr	0 (999 31y 240		720
att Ile	aga Arg	tat Ty:	c G	In 1	cac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	ca G1 25	n G	ga ly	tgg Trp	aaa Lys	gga Gly	tc Se 25	r I	cca Pro		768
ALG	ata Ile	Pne	26	in S	er :	Ser	Met	Thr	Lys 265	Il	e L	eu (Glu	Pro	Phe 270	Ar	g I	ys		816
	aat Asn	275	, G.	LU I	TG ,	AGT	iie	280	Gin	ТУ	r V	al A	Asp	Asp 285	Leu	Ty	r V	'al		864
gga Gly	tct Ser 290	gac Asp	Le	a g u G	aa a lu 1	тте	gga Gly 295	cag Gln	cat His	aga Arg	a a	hr I	aaa Lys	ata Ile	gag Glu	gaa Glu	a y	ta aa		912

-148-

305 310 315 320	960
cag aaa gaa cct cca ttt ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata cag ctg cca gaa aag gaa agc tgg act Lys Trp Thr Val Gln Pro Ile Gln Leu Pro Glu Lys Glu Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa ttg aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat gca ggg Ile Tyr Ala Gly 370	1116
<210> 82 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116)	
<223> Portion of HIV Reverse Transcriptase	
	48
<223> Portion of HIV Reverse Transcriptase <400> 82 cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly	4 8 96
<pre><223> Portion of HIV Reverse Transcriptase <400> 82 cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly 1</pre>	
<pre><223> Portion of HIV Reverse Transcriptase <400> 82 cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly 1</pre>	96
<pre><223> Portion of HIV Reverse Transcriptase <400> 82 cct cag atc act ctt tgg caa cga ccc ctc gtc aca gta aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly 1</pre>	96 144

-149-

	tta Lei	a aa 1 As:	t tt n Ph	t cc e Pr 10	o Il	t ag e Se:	t cci	t at	t gaa e Glu 109	ı Thi	gta Val	a cca l Pro	a gta	a aaa l Lys	Le	a aag u Lys	336
	Pro	gg Gl	a at y Me 11	t As	t gg p Gl	c cca y Pro	a aaa o Lys	a gti s Val	l Lys	a caa s Glr	tgg Trp	g cca p Pro	tto Lev 125	ı Thi	a gaa Gli	a gaa ı Glu	384
	aaa Lys	ata Ile 130	E Ly	a gc s Al	a tta a Lei	a gta ı Val	a gaa l Glu 135	ı Ile	tgt Cys	aca Thr	gaa Glu	a ato Met 140	: Glı	a aag 1 Lys	g gaa Glu	a ggg u Gly	432
	aaa Lys 145	Ile	tc: Se:	a aa r Ly	a att	999 Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
	gcc Ala	ata Ile	aaa Lys	a aag s Lys	g aaa s Lys 165	: Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	Arg	aag Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
	aga Arg	gaa Glu	ctt Lei	aay Asi 180	ı Lys	aag Lys	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Leu	gga Gly	576
	ata Ile	cca Pro	Cat His 195	Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
	gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gam Xaa	ttc Phe	agg Arg	672
i	aar Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
:	att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
į	gca Ala	ata Ile	ttt Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
G	aa In	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tay Tyr 280	cag Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
9	TÀ	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	ggr Xaa 295	aag Lys	cac His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gag Glu	cta Leu	912
A	ga xg 05	cag Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
G	ag ln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctk Xaa	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008

-150-

aaa Lys	a tg	g ac p Th	a gt r Va 34	l Gl	g cc	t ata	a aa e Ly	a ctes S Les 34	u Pr	a ga o Gli	a aa u Ly	a ga s As	c ag p Se 35	r Tr	g act p Thr	1056
gty Xaa	/ aa a As:	t ga n As 35	p Il	a caq e Gli	g aaq n Lys	g tta s Lei	a gte 1 Va. 360	l Gl	a aa y Ly:	a tt: s Xaa	r aat a Ast	t tgg n Trj 369	o Al	c ag a Se:	t cag r Gln	1104
att		t gca r Ala														1116
<21 <21	0> 8 1> 1 2> I 3> I	1116	ı Imn	nunoc	lific	:ienc	y Vi	.rus	(HIV	7)						
<22	1> (2> (DS (0)														
<22	1> C 2> (3> F	DS (298) Orti	(on c	1116 f HI) V Re	vers	e Tr	ansc	ript	ase						
cct	0> 8 cag Gln	ato	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	cca Pro	ctc Leu 10	Val	gca Ala	ata Ile	aag Lys	ata Ile 15	ggg	48
gly aaa	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta Leu	gaa Glu	gac Asp 35	atg Met	agt Ser	tta Leu	cca Pro	gga Gly 40	aaa Lys	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly 999	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	caa Gln	gta Val	ccc Pro	ata Ile	192
gaa Glu 65	atc Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca (gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384

-151-

aaa Lys	a ata s Ile 130	: Lys	a gca s Ala	a tta a Lei	a gta ı Val	gaa Glu 135	ı Ile	tgt Cys	aca Thr	gaa Glu	ato Met	: Glı	a aag 1 Lys	g gaa s Glu	a gga a Gly	432
aaa Lys 145	: Ile	tca Ser	aaa Lys	att Ile	ggg Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tat Tyr 155	Asr	act Thr	cca Pro	gta Val	ttt Phe 160	480
gco					Asp					Arg					ttt Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gat Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	Caa Gln 190	Leu	gga Gly	576
													Thr		ctg Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	ccc Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	ccc Pro	agg Arg 240	720
gtt Val	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gta Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
		ttc Phe														816
caa Gln	aac Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gac Asp 285	tta Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gag Glu	ata Ile	gga Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gac Asp	cta Leu	912
aga Arg 305	gca Ala	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	ggg Gly	Phe	Thr	aca Thr 315	Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	ccc Pro	cca Pro 325	ttt Phe	ctc Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gwg Xaa	cta Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aaa Lys	tta Leu	gta Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104

-152-

att tat cca ggg Ile Tyr Pro Gly 370	1116
<210> 84 <211> 1116 _<212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 84 cct caa atc act ctt tgg caa cga ccc att gtc aca ata aaa gta ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Val Gly 1 5 10 15</pre>	48
ggg caa cta atg gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Met Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gac ata aat ttg cca gga aga tgg aaa cca aaa ata ata ggg Leu Glu Asp Ile Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly 35 40 45	144
gga att ggt ggt ttt gtc aaa gtg aga cag tat gat cag gta ccc ata Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Val Pro Ile 50 55 60	192
gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240
cct acc aac gta gtt gga aga aat ctg atg act cag att ggc tgc acy Pro Thr Asn Val Val Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Xaa 85 90 95	288
tta aat ttt cct att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg acg gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gat gga Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Asp Gly 130 . 135 140	432
aaa att tca aaa att ggg cct gaa aat cca tat aat act cca ata ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe 145 150 155 160	480

-153-

go Al	ec at la II	ta a le L	ag a	aaa Lys	Lys 165	ASI	agi n Se:	ga Asj	t aaa o Lys	tg Trj	p Ar	a aa g Ly:	a tta s Lei	a gta 1 Val	a gat L Asp 175	ttc Phe	528
ag Ar	ga ga g G]	aa c lu L	eu A	aat Asn 180	aar Lys	aga Arg	act Thi	caa Glr	a gad n Asp 185	Phe	c tgg ≥ Trp	g gaa o Glu	a gtt 1 Val	caa Glr 190	Leu	gga Gly	576
at Il	a co e Pr	ю н:	at c is P 95	ect Pro	gca Ala	Gly	tta Leu	aaa Lys 200	: Lys	aat Asr	aaa Lys	tca Ser	a gta Val 205	Thr	gta Val	ctg Leu	624
ga As	t at p Il 21	e GI	y A	gat Asp	gca Ala	tat Tyr	Phe 215	Ser	att : Ile	Pro	tta Leu	gat Asp 220	. Lys	gac Asp	ttt Phe	agg Arg	672
aa Ly: 22:	з ту	t ac r Th	t g r A	gca lla	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	. Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
gt: Va:	t ag l Ar	a ta g Ty	t c r G	ın	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	Gly	tgg Trp	aag Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	a ata	a tt e Ph	e G	aa ln 60	agc Ser	agc Ser	atg Met	acc Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
Glr	g aat 1 Asi	cc Pr 27	O A	ac sp	ata Ile	gtt Val	atc Ile	tgc Cys 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
GIY	290	AS ₁	5 те	eu (GIU	11e	G1y 295	GIn	His	Arg	Thr	100	ata Ile	Glu	Glu	Xaa	912
305	ASI	Ada	a Tre	eu 1	rrp	Lys 310	Trp	Gly	Phe	Tyr	Thr 315	Pro	gac Asp	Lys	Lys	Tyr 320	960
GIII	гÀг	GIL	l PI	3	325	Pne	Leu	Trp	Met	Gly 330	Tyr	Glu	ctc Leu	His	Pro 335	Asp	1008
aaa Lys	tgg Trp	aca Thr	yt Va 34	ידי	ag (Sln)	ecc Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	TT	a c e G	ag a	aag Lys :	Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile	tat Tyr 370	cca Pro	99 G1	g Y													1116

<210> 85 <211> 1116 WO 01/35316 PCT/US00/30863

-154-

<212 <213			ıImm	unod	lific	ienc	y Vi	.rus	(HIV	')						
	L> C 2> (0)	.(29 rote													
	!> (298)	(on o			vers	e Tr	ansc	ript	ase						
<400 cct Pro 1	caa	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	ctc Leu 10	gtc Val	aca Thr	ata	aaa Lys	gta Val 15	ely aaa	48
										gga Gly					gta Val	96
tta Leu	gaa Glu	gaa Glu 35	atg Met	aat Asn	ttg Leu	cca Pro	999 Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	ggg Gly	144
gga Gly	att Ile 50	gga Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	gta Val	agc Ser	ata Ile	192
gaa Glu 65																240
cct o	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
tta a Leu A																336
cca (Pro (384
aaa a Lys]																432
aaa a Lys I 145	att [le	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc a Ala I	ita []e	aar Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aag Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga g Arg G	gaa Slu	ctt Leu	aat Asn 180	aaa Lys	ara Xaa	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576

-155-

	ata Ile	cca Pro	cat His 195	Pro	gca Ala	Gly	tta Leu	aaa Lys 200	Lys	aam Xaa	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	ctg Leu	624
	gay Asp	gtg Val 210	Gly	gat Asp	gcr Xaa	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ссу Хаа	tta Leu	gay Asp 220	aaa Lys	gay Asp	ttc Phe	agg Arg	672
	aag Lys 225	tac Tyr	aca Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aac Asn 235	aat Asn	gag Glu	rca Xaa	cca Pro	999 Gly 240	720
	att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	car Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
;	gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aar Lys	816
7	maa Kaa	aat Asn	cca Pro 275	gac Asp	ata Ile	gty Xaa	atc Ile	tay Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttr Xaa	tat Tyr	gta Val	864
ć	gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gga Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
F	aga Arg 805	caa Gln	cat His	ctg Leu	ttg Leu	cag Gln 310	tgg Trp	Gly 999	tta Leu	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
G	ag Sln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	999 Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	ccg Pro 335	gat Asp	1008
a I	aa ys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	wtg Xaa	ctg Leu 345	cca Pro	gac Asp	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
9 X	aa i	Asn	gac Asp 355	ata Ile	cag Gln	aar Lys	tta Leu	gta Val 360	gga Gly	aaa Lys	ttg Leu	Asn	tgg Trp 365	gcg Ala	agt Ser	cag Gln	1104
	le :		cca Pro									•					1116
<.	211: 212:	> 86 > 11 > DN > Hu	A	Immu	nodi	fici	ency	Vir	na (HIV)							
<: <:	222>	> CD:	S) V Pro														

-156-

<221> CDS <222> (298)...(1116) <223> Portion of HIV Reverse Transcriptase <400> 86 cct caa atc act ctt tgg caa cga ccc atc gtc aca gta aag ata ggg 48 Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Val Lys Ile Gly ggg cac aca acg gaa gct cta tta gat aca gga gca gat gat aca gta Gly His Thr Thr Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 96 tta gaa gaa atg aat ttg cca ggg aga tgg aaa cca aaa atg ata gga 144 Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly gga att gga ggt ttt atc aaa gta aga cag tat gag cag gta ccc ata Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Glu Gln Val Pro Ile 192 gaa ttc tgt gga cat aaa act gta ggt aca gta tta ata gga cct aca Glu Phe Cys Gly His Lys Thr Val Gly Thr Val Leu Ile Gly Pro Thr 240 cct gtc aac ata att gga aga aat ctg atg act cag att ggt tgt act Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr 288 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 336 100 105 384 cca gga atg gat ggg ccc aaa gtt aaa cca tgg cca ttg aca gaa aga Pro Gly Met Asp Gly Pro Lys Val Lys Pro Trp Pro Leu Thr Glu Arg 120 432 aaa aat aaa gca tta gta gaa att tgt tcc gaa atg gaa aaa gga agg Lys Asn Lys Ala Leu Val Glu Ile Cys Ser Glu Met Glu Lys Gly Arg 480 aaa att tca aaa att ggg cct gag aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe gcc ata aag aaa aag aac agt act aga tgg aga aaa tta gta gat ttc 528 Ala Ile Lys Lys Lys Asn Ser Thr Arg Trp Arg Lys Leu Val Asp Phe 165 576 aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt cag tta gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 624 ata cca cat ccc gca ggg tta aaa aag aac aaa tca gta aca gta ctg Ile Pro His Pro Ala Gly Leu Lys Lys Asn Lys Ser Val Thr Val Leu 200 gat gta ggt gat gca tat ttt tca gtt ccc tta gat gaa gaa ttc agg Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Glu Phe Arg 672 210

WO 01/35316 PCT/US00/30863

-157-

aag Lys 225	tat Tyr	act Thr	gca Ala	ttc Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gaa Glu	aca Thr	cca Pro	999 Gly 240	720
		tat Tyr														768
		ttc Phe														816
		cca Pro 275														864
		gat Asp														912
		cat His														960
		gaa Glu														1008
		aca Thr														1056
		gat Asp 355														1104
		cca Pro														1116
<211 <212	> 87 > 11 > DN > Hu	16	Immu	nodi	fici	ency	. Vir	rus (HIV)							
<222	> CD > (0	S) V Pr														
<221 <222 <223	> (2	98).		-	Rev	erse	Tra	nscr	ipta	.se						
<400 cct Pro	cag	atc														48

-158-

999 Gl ₃	g caa / Glr	t cta Leu	aag Lys 20	Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gta Val	96
			Met					Arg	tgg Trp							144
		Gly							cag Gln			Gln				192
	Ile					Āla			aca Thr							240
									ttg Leu 90							288
									act Thr							336
									caa Gln							384
									aca Thr							432
									cca Pro							480
									tgg Trp 170							528
									ttc Phe							576
									aaa Lys							624
									ccc Pro							672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
gtt Val	aga Arg	tat Tyr	car Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	Pro	cag Gln 250	gga Gly	tgg Trp	aag Lys	gga Gly	tca Ser 255	cca Pro	768

-159-

gca Ala	ata Ile	ttc Phe	caa Gln 260	agc Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gat Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	Gln	tac Tyr	atg Met	gat Asp	gac Asp 285	ttr Xaa	tat Tyr	gta Val	864
gga Gly	ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	car Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ttg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	tta Leu	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gl _i n	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gat Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
	tat Tyr 370															1116
<213 <213	0> 88 L> 11 2> DN 3> Hu	16 A	Immu	nodi	fici	ency	, Vir	rus (ĤIV)			-				
<222)> L> CD !> (0 !> HI)														
<222	> CD > (2 > Po	98).			Rev	erse	Tra	nscr	ipta	se						
cct	> 88 cag Gln	atc (act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	atc Ile 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	gly aaa	48
Gly	caa (Gln)	cta a Leu A	agg Arg : 20	raa (Kaa .	gct (Ala :	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta Leu	gaa q Glu <i>l</i>	gac a Asp : 35	ata (gaa Glu	ttg (Leu 1	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 1 45	atg Met	ata Ile	ggg ggg	144

-160-

	gga Gl _y	a ati	e Gl	a gg y Gl	t tti y Pho	t gto e Val	aaa L Lys 59	va.	a aga l Arg	a caa g Gli	a tat 1 Tyr	t gat r Asp 60	Glr	g ata 1 Ile	e cco	ata Ile	19:
	gaa Glu 65	ı Ile	tg Cy	t gg s Gl	a cat y His	t aaa S Lys 70	: Val	ata l Ile	a ggt e Gly	aca Thi	gta Val	Leu	a gta ı Val	a gga L Gly	cct Pro	aca Thr 80	240
	Pro	gcc Ala	a ac	c ata	a att e Ile 85	e Gly	aga Arg	aat JAsr	ctg Leu	ato Met 90	Thr	cag Glr	g ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
	tta Leu	aat Asn	ttt Phe	ccc Pro 100) Il∈	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336
	cca Pro	gga Gly	ato Met	Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	. Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	Thr	aaa Lys	gaa Glu	384
	aaa Lys	ata Ile 130	Glu	gca Ala	tta Leu	atr Xaa	gaa Glu 135	Ile	tgt Cys	gma Xaa	ttt Phe	ttg Leu 140	Glu	aag Lys	gaa Glu	gga Gly	432
	aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	ccg Pro	tac Tyr 155	aac Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
	gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gga Gly	ggt Gly	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	ata Ile	gta Val	gat Asp 175	ttc Phe	528
,	aga Arg	gaa Glu	ctt Leu	aat Asn 180	aaa Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
	ata Ile	cca Pro	cat His 195	ccc Pro	gcg Ala	gly aaa	tta Leu	aaa Lys 200	aag Lys	aay Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
9	gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	att Ile	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gaa Glu	ctc Leu	agg Arg	672
1	aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	act Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
3	att [le	aga Arg	tac Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
Ē	gca Ala	ata Ile	ttt Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	ccc Pro	ttt Phe 270	aga Arg	aag Lys	816
G	aa Sln	Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	Ile	twt Xaa 280	caw Xaa	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864

WO 01/35316 PCT/US00/30863

-161-

Gly Ser Asp Leu Glu Ile Gly Lys His Arg Glu Lys Ile Glu Glu Leu 290 295 300	912
aga caa cat ctg tgg aag tgg gga ttt tac aca cca gac gaa aaa cat Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Glu Lys His 305 310 315 320	960
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat ctt gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Leu Asp 325 330 335	1008
aaa tgg aca gta cag cct ata gtg ctg cca gaa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat gca ggg Ile Tyr Ala Gly 370	1116
<210> 89 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220>	
<221> CDS <222> (0)(297) <223> HIV Protease	
<222> (0)(297)	
<222> (0)(297) <223> HIV Protease <221> CDS <222> (298)(1116)	48
<pre><222> (0)(297) <223> HIV Protease <221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase <400> 89 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly</pre>	48 96
<pre><222> (0)(297) <223> HIV Protease <221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase <400> 89 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1</pre>	
<pre><222> (0)(297) <223> HIV Protease <221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase <400> 89 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1</pre>	96

-162-

					Gly					Thr					act	288
									Pro						aag Lys	336
								Lys							gaa Glu	384
		Lys													Gly 999	432
	Ile											act Thr				480
												tta Leu				528
												gtt Val				576
												gta Val 205				624
												aaa Lys				672
												gag Glu				720
												aaa Lys				768
												cct Pro				816
cag Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gac Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	gaa Glu	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960

This page is not part of the pamphlet!

WO 01-35316 5/5

Date: 17 may 2001

Destination: Agent

WO 01/35316 PCT/US00/30863

-163-

Gli	g aaa Lys	gaa Glu	cct Pro	cca Pro 325	Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ct <i>c</i> Leu	cat His	cct Pro 335	gat Asp	1008
								ctg Leu 345								1056
			Ile					gga Gly								1104
		gca Ala						-								1116
<21 <21	0 > 9 1 > 1 2 > D 3 > H	116 NA	Imm	ınod:	ific:	iency	y Vi:	rus	(HIV))						
<22	1> C 2> (DS 0) IV P:														
	1> C 2> ((:	1316	1											
	3> P					erse	Tra	ansci	ipta	ase						
<22 <40 cct	0> 9 cag	ortie 0 atc	on of	E HIV	/ Rev	caa	cga	ccc Pro	aty	gtc						48
<22 <40 cct Pro 1	0> 9 cag Gln cag	ortic 0 atc Ile cta	act Thr	ctt Leu 5	tgg Trp gct	caa Gln yta	cga Arg tta	ccc	aty Xaa 10	gtc Val gga	Thr	Ile gat	Lys gat	Val 15 aca	Gly gta	48 96
<22 <40 cct Pro 1 gga Gly	0> 9 cag Gln cag Gln	ortico 0 atc Ile cta Leu	act Thr aag Lys 20	ctt Leu 5 gaa Glu	tgg Trp gct Ala	caa Gln yta Xaa cca	cga Arg tta Leu	ccc Pro gat Asp	aty Xaa 10 aca Thr	gtc Val gga Gly	Thr gca Ala cca	Ile gat Asp	Lys gat Asp 30 ata	Val 15 aca Thr	Gly gta Val 999	
<22 <40 cct Pro 1 gga Gly tta Leu gga	0> 9 cag Gln cag Gln gaa Glu	orticortico atc Ile cta Leu gaa Glu 35 gga Gly	act Thr aag Lys 20 atg Met	ctt Leu 5 gaa Glu aac Asn ttt	tgg Trp gct Ala ttg Leu	caa Gln yta Xaa cca Pro aga Arg	cga Arg tta Leu gga Gly 40 gta Val	ccc Pro gat Asp 25	aty Xaa 10 aca Thr tgg Trp	gtc Val gga Gly aaa Lys tat	Thr gca Ala cca Pro gat Asp	gat Asp aaa Lys 45	gat Asp 30 ata Ile	Val 15 aca Thr ata Ile	Gly gta Val ggg Gly	96
<22 <40 cct Pro 1 gga Gly tta Leu gga Gly	0> 9 cag Gln cag Gln gaa Glu att Ile 50 att	orticolor	act Thr aag Lys 20 atg Met ggt Gly	ctt Leu 5 gaa Glu aac Asn ttt Phe	tgg Trp gct Ala ttg Leu gtc Val	caa Gln yta Xaa cca Pro aga Arg 55	cga Arg tta Leu gga Gly 40 gta Val	ccc Pro gat Asp 25 aaa Lys	aty Xaa 10 aca Thr tgg Trp caa Gln	gtc Val gga Gly aaa Lys tat Tyr	Thr gca Ala cca Pro gat Asp 60 tta	gat Asp aaa Lys 45 cag Gln	Lys gat Asp 30 ata Ile gta Val	Val 15 aca Thr ata Ile cct Pro	Gly gta Val ggg Gly gta Val aca	96 144
<22 <40 cct Pro 1 gga Gly tta Leu gga Gly gaa Glu 65 cct	0> 9 cag Gln cag Gln gaa Glu att Ile 50 att Ile	orticonticonticonticonticonticonticonticon	act Thr aag Lys 20 atg Met ggt Gly gga Gly	ctt Leu 5 gaa Glu aac Asn ttt Phe cat	tgg Trp gct Ala ttg Leu gtc Val aaa Lys 70 gga	caa Gln yta Xaa cca Pro aga Arg 55 gct Ala	cga Arg tta Leu gga Gly 40 gta Val ata Ile	ccc Pro gat Asp 25 aaa Lys aga Arg	aty Xaa 10 aca Thr tgg Trp caa Gln tca Ser	gtc Val gga Gly aaa Lys tat Tyr gta Val 75 act	Thr gca Ala cca Pro gat Asp 60 tta Leu cag	gat Asp aaa Lys 45 cag Gln gta Val	Lys gat Asp 30 ata Ile gta Val gga Gly	Val 15 aca Thr ata Ile cct Pro	gta Val ggg Gly gta Val aca Thr 80	96 144 192

-164-

			Asp					Lys							gaa Glu	384
		Lys													gga Gly	432
aag Lys 145	Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
					aac Asn											5.28
					aga Arg											576
					ggg Gly											624
					tat Tyr											672
					acc Thr 230											720
att Ile	aga Arg	tat Tyr	car Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
					agc Ser											816
					gtt Val											864
	Ser	Āsp	Leu	Glu	ata Ile	Glu	Gln		Arg	Thr	Lys	Ile				912
					agg Arg 310											960
					ttc Phe											1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cat His	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056

-165-

gtc aat gac ata cag aag tta gtg gga aaa ttr aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Xaa Asn Trp Ala Ser Gln 355 360 365	1104
att tat gca ggg Ile Tyr Ala Gly 370	1116
<210> 91 <211> 1115 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1115) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 91 cct cag atc act ctt tgg caa cga ccc ctt gtc aca gta aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Val Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta ata gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Ile Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
ttg gaa gaa atg aat ttg cca ggg aga tgg aaa cca aaa ata ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Ile Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ccc ata Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile 50 55 60	192
gaa atc tgt gga cat aaa gtt ata rgt cca gta tta ata gga cct aca Glu Ile Cys Gly His Lys Val Ile Xaa Pro Val Leu Ile Gly Pro Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ttg atg act cag att ggc tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Met Thr Gln Ile Gly Cys Thr 85 90 95	288
tta aat ttt ccc atc agt cct att raa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Xaa Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aag gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135 140	432

-166-

aaa Lys 145	s Ile	tca Ser	aaa Lys	att	999 Gly 150	Pro	gaa Glu	aac Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
			aaa Lys		Asn					Arg						528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cct Pro	gga Gly	Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	cct Pro	cta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
	Tyr		gca Ala													720
gtt Val	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tcg Ser 255	cca Pro	768
gca Ala	ata Ile	ttt Phe	cag Gln 260	gct Ala	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	ccg Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	cta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ttg Leu	ttg Leu	aaa Lys 310	tgg Trp	gga Gly	ttt Phe	atc Ile	aca Thr 315	cca Pro	gat Asp	gaa Glu	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	Phe	Leu	\mathtt{Trp}	atg Met	Gly	Tyr	Glu	Leu	His	cct Pro 335	gat Asp	1008
aag Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gta Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aaa Lys	Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile	tat Tyr 370	gca Ala	9 9													1115

-167-

<2: <2: <2: <2: <2: <2: <2:	20> 21> (22>	1116 DNA Humar CDS (0)	ı Imr	97)	lific	cieno	cy Vi	.rus	(HIV	7)						
<.22		(298)	(.on c			vers	e Tr	ansc	ript	ase						
cct	00> 9 cag Glr	ato	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc	ctc Leu 10	Val	aca Thr	ata Ile	aag Lys	ata Ile 15	Gly	48
ggg Gly	cag Gln	cta Leu	aag Lys 20	Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	Thr	gta Val	96
tta Leu	gaa Glu	gac Asp 35	Ile	aac Asn	ttg Leu	cca Pro	gga Gly 40	aaa Lys	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly ggg	144
gga Gly	att Ile 50	Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gag Glu 60	cag Gln	gta Val	ccc Pro	ata Ile	192
gaa Glu 65	Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	act Thr	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	atg Met 90	act Thr	cag Gln	att Ile	ggg Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gj ^a aaa	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aag Lys 165	aac Asn	agt Ser	act Thr	aga Arg	tgg Trp 170	aga Arg	aaa Lys	gta Val	gta Val	gat Asp 175	ttc Phe	528

-168-

								gac Asp 185								576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	999 Gly	tta Leu	aaa Lys 200	aag Lys	aac Asn	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
								gtt Val								672
_			_			_		agt Ser	_				_			720
								cca Pro								768
								aaa Lys 265								816
								caa Gln								864
								cat His								912
								ttt Phe								960
								atg Met								1008
								ctg Leu 345								1056
								gga Gly								1104
		gca Ala														1116

<210> 93 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)

WO 01/35316 PCT/US00/30863

-169-

<221> CDS <222> (0)...(297) <223> HIV Protease <221> CDS <222> (298) ... (1116) <223> Portion of HIV Reverse Transcriptase <400> 93 cct cag atc act ctt tgg caa cga ccc atc gtc aca ata aag ata gga Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly ggg cag cta aag gaa gct cta ata gat aca gga gca gat gat aca gta 96 Gly Gln Leu Lys Glu Ala Leu Ile Asp Thr Gly Ala Asp Asp Thr Val tta gaa gaa atg aat tta cca gga aga tgg aca cca aaa ata ata ggg 144 Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Thr Pro Lys Ile Ile Gly gga att gga ggt ttt gtc aga gta aga cag tat gaa cag ata ccc gta 192 Gly Ile Gly Gly Phe Val Arg Val Arg Gln Tyr Glu Gln Ile Pro Val gaa atc tgc ggg cat aaa gct gta ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr cct gcc aac ata att gga aga aat ctg ttg act cag att ggc tgt act 288 Pro Ala Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr tta aat ttt ccc att agt cct att gat act gta cca gta aaa tta aag 336 Leu Asn Phe Pro Ile Ser Pro Ile Asp Thr Val Pro Val Lys Leu Lys 105 cca gga atg gat ggc cca ara gtt aaa caa tgg cca ttg aca gaa gag Pro Gly Met Asp Gly Pro Xaa Val Lys Gln Trp Pro Leu Thr Glu Glu 384 432 aaa ata aaa gca tta gta gaa att tgt aca gaa ctg gaa aag gam gga Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Xaa Gly 135 480 aaa att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 150 gct ata aag aaa aaa gac agt act aaa tgg aga aaa gta gta gat ttc 528 Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Val Val Asp Phe aga gaa ctt aat aaa aga act caa gac ttc tgg gaa gtt caa tta gga 576 Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 180 ata cca cat cct gca ggg ata maa aag aac aaa tca gta aca gta ytg Ile Pro His Pro Ala Gly Ile Xaa Lys Asn Lys Ser Val Thr Val Xaa 624

-170-

					tat Tyr									672
					acc Thr 230									720
					aat Asn									768
					agc Ser									816
					rtt Xaa									864
					ata Ile									912
					aag Lys 310									960
					ttc Phe									1008
					cct Pro									1056
					aag Lys									1104
Asn		gca Ala											;	1116
<210 <211 <212 <213	> 11 > DN	16 A	Immu	nodi	fici	ency	Vir	us (HIV)					
<220 <221 <222 <223	> CD > (0)												
<221 <222 <223	> (2	98).			Rev	erse	Tra	nscr	ipta	se				

-171-

	00>															
ec Pr 1	t ca o Gl	g at n Il	c ac e Th	t ct r Le 5	t tg u Tr	g caa p Gli	a cga	g Pro	c cto Lev	ı Val	c aca	a ata r Ile	a aag e Lys	g ata s Ile 1	a ggg e Gly 5	48
99 G1	g ca y Gl:	a ct n Le	u Il	a ga e Gl 0	g gc u Ala	t cta a Leu	tto Lev	g gat 1 Asp 25	Thr	a gga Gly	a gca / Ala	a gat a Asp	gat Asp 30	Th:	a gta r Val	96
tt: Le:	a gaa 1 Glu	a ga u Gl 3.	u Me	g ga t As	t ttg p Lei	g cca ı Pro	gga Gly 40	' Arg	tgg Trp	aaa Lys	cca Pro	a aaa D Lys 45	: Ile	ata : Ile	a ggg e Gly	144
gga Gly	a att / Ile 50	≥ G1;	a gg y Gl	t tg: y Tr]	g ato p Ile	aaa Lys 55	Val	aga Arg	caa Gln	tat Tyr	gat Asp 60	Glr	ata Ile	ccc Pro	ata Ile	192
gaa Glu 65	ı Ile	tgt Cys	t gga s Gly	a cat y His	aaa Lys 70	Val	ata Ile	agt Ser	aca Thr	gta Val 75	Leu	gta Val	gga Gly	cct	aca Thr 80	240
cca	gto Val	aac Asr	gta n Val	a att l Ile 89	Gly	aga Arg	aat Asn	ctg Leu	atg Met 90	act Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	Thr	288
tta Leu	aat Asn	ttt Phe	Pro) Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met	Asp	ggc Gly	cca Pro	aga Arg	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aag Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gat Asp	Gly 999	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	gta Val	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	Asp	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	Gly ggg	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	ggg Gly	tta Leu	cca Pro 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
aaa Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	Ile .	aat Asn . 235	aat Asn	gag Glu	aca Thr	Pro	gga Gly 240	720

-172-

gti Val	t aga l Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctc Leu	cca Pro	cag Gln 250	Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro		768
gca Ala	a ata a Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	acc Thr	aaa Lys 265	Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys		816
cag Gln	g aat Asn	cca Pro 275	aac Asn	ata Ile	ctt Leu	att Ile	tgt Cys 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val		864
gga Gly	Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	gaa Glu 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu		912
aga Arg 305	caa Gln	cat His	ctg Leu	tgg Trp	aga Arg 310	tgg Trp	gly aaa	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gat Asp	aaa Lys	aaa Lys	cat His 320		960
cag Gln	aag Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	:	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gag Glu	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	:	1056
gtc Val	aat Asn	gat Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agy Xaa	cag Gln	3	1104
att Ile	tat Tyr 370	gca Ala	gly 999													נ	.116
<211 <212)> 95 .> 11 !> DN: !> Hu	A	Immu	nodi	fici	ency	Vir	us ()	HIV)								·
<222	> CD: > (0 > HI	·															
<222	> CD: > (2: > Po:	98).	(1] n of	l16) HIV	Reve	erse	Trai	nscr:	ipta:	se							
<400 cct Pro	cag a	atc a [le]	et c Thr I	tt t eu 1 5	tp 6	aa c Sln <i>P</i>	ga d Arg F	ecc o	ctc c Leu V	gtc a Val 1	aca a Thr i	ata a Ile 1	aag a Lys 1	ata (le (gly ggg		48
gjå (aaa (caa c Gln I	ta a eu I	ag g ys G 20	aa g lu A	ct c la I	ta t eu I	ta g eu A	at a Asp 1 25	nca g	gga g Sly A	gca g Nla A	gat o Asp A	gat a Asp 1	aca c	gta Val		96

-173-

			Met					Arg							gj aaa	144
		Gly						aga Arg								192
	Ile							ggt Gly								240
	_							ttg Leu	_		_			_		2.8.8
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
								aaa Lys								384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
	Ile							aat Asn								480
								aaa Lys								528
								gac Asp 185								576
								aag Lys								624
								gtt Val								672
aag Lys 225	tac Tyr	act Thr	gca Ala	ttt Phe	act Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
								aaa Lys 265								816

-174-

ca Gl	a aa .n As	t cc n Pr 27	o Glu	ata Ile	gtt Val	atc Ile	tat Tyr 280	Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	Let	g tai 1 Tyi	gta Val	864
99 G1	y Se 29	r Ası	tta Leu	gaa Glu	ata Ile	gaa Glu 295	cag Gln	cat His	aga Arg	ata Ile	aaa Lys 300	$Il\epsilon$	gag Glu	ggaa i Gli	a ctg Leu	912
ag Ar 30	g Hi	c cat s His	ctg Leu	ttg Leu	aaa Lys 310	tgg Trp	gga Gly	ttt Phe	wmc Xaa	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
ca Gl:	g aaa n Lys	a gaa s Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	Asp	1008
aa: Ly:	a tgg s Trp	g aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aar Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gto Va	c aat l Asr	gac Asp 355	ITe	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
	tac Tyr 370	Pro														1116
<21 <21	.0> 9 .1> 1 .2> D .3> H	116 NA	Immu	ınodi	fici	iency	Vir	rus (HIV)							
<22	1> C 2> (0)	. (297 rotea													
<22		298)	(1 on of		' Rev	erse	Tra	nscr	ipta	se						
cct	0> 90 caa Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln 2	cga Arg	ccc Pro	aat Asn 10	gtc (Val (aca Thr	gta Val	aag Lys	ata Ile 15	ggr Xaa	48
ggg Gly	caa Gln	cta Leu	agg Arg 20	gaa Glu	gct Ala	cta i Leu i	tta Leu .	gat Asp 25	aca (Thr (gga g Gly i	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta Leu	gaa Glu	gaa Glu 35	ata Ile i	aat : Asn :	ttg Leu	cca g Pro (ga Gly 40	aga (Arg (rgg a	aaa d Lys 1	cca Pro	aaa Lys 45	atg Met	ata Ile	G1y ggg	144
gga																

-175-

ga Gl 6	u 11	c to	gt g ys G	ga ca ly Hi	s Ly	a go s Al	t at a Il	a gg e Gl	t ac y Th	a gt r Va 7	l Le	a gt u Va	a gg l Gl	a cc y Pr	c aca o Thr	•
cc Pr	t gt o Va	c aa l As	ac at	re II	t gg e Gl 5	a ag y Ar	a aa g As:	t cto n Leo	g ttg u Lei 90	ı Th:	t ca r Gl:	g at n Il	t gg e Gl	t tg y Cy 9	c act s Thr 5	288
tt: Le:	a aa u As	t tt n Ph	t co le Pr 10	o II	t ag e Se	t cc r Pr	t at	t gaa e Glu 109	ı Thi	gta Val	a cca l Pro	a gt.	a ara l Xaa	a Le	a aag u Lys	336
Pro	a gg: o Xa	r at a Me 11	t As	t gg	c cc y Pro	a aaa o Lys	a gti s Val 120	l Lys	a caa s Glr	tgg Trp	g cca Pro	tto Lev 125	ı Thi	a ga r Gli	a gaa u Glu	384
aaa Lys	a ata 5 Ile 130	s nà	a go s Al	a tta a Lei	a gta ı Val	a gaa l Glu 135	ı ⊤T€	tgt Cys	aca Thr	gaa Glu	ato Met 140	: Gli	a aag 1 Lys	g gaa s Glu	a ggg	432
aaa Lys 145	: TT6	tc Se	a aa r Ly	a att s Ile	ggg Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	cca Pro	ata O Ile	ttt Phe 160	480
gcc Ala	ata Ile	aa _: Ly:	g aa s Ly	a aaa s Lys 169	Asp	ggt Gly	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	gta Val	gta Val	gat Asp 175	ttc Phe	528
agg Arg	gaa Glu	cto Le	aat 1 Ast 180	r TAE	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	ggm Xaa	576
ata Ile	cca Pro	cat His 199	Pro	gca Ala	ggg	ttg Leu	aaa Lys .200	Lys	aaa Lys	aaa Lys	tca Ser	gtr Xaa 205	aca Thr	gta Val	ctg Leu	624
gat As p	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gaa Glu	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
atc Ile	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aag Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttt Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtc Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
Gry	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912

-176-

	a ca g Gl: 5	a ca n Hi	t ctg s Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	tt: Phe	t aco	c acar Thi	Pro	a ga	c aa p Ly	a aa s Ly	a ca s Hi 32	.s
Ca Gl:	g aaa n Lys	a gaq s Gli	g cct u Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	ato Met	g ggt E Gl _y 330	/ Ty:	gaa Glu	a cto 1 Lei	c ca ı Hi	t cc s Pro 33!	o As	t 1008 p
aa: Ly:	a tgg s Trp	g aca o Thi	a gta Val 340	cag Gln	cgt Arg	ata Ile	gag Glu	Cto Lev 345	ı Pro	a gaa o Glu	aag Lys	g gag Glu	g ag 1 Se: 350	r Trī	g ac o Th	t 1056 r
gto Val	aat L Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	Ala	a agt a Ser	caa Gli	a 1104 n
atw Xaa	tac Tyr 370	Pro	Gly													1116
<21 <21	0 > 9 1 > 1 2 > D 3 > H	116 NA	Immu	nodi	fici	ency	Viz	rus	(HIV)						
<22 <22 <22	0> 1> Cl 2> ((DS 0)	.(297 rotea)		_										
<22	1> CI 2> (2 3> Po	298).	(1 on of	116) HIV	Rev	erse	Tra	nsci	ripta	ase						
<22: <22: <40: cct	2> (2 3> Po 0> 97 caa	298). ortic 7 atc	(1 on of act (HIV	caa o	caa (cga	ccc	ctc	atc	aaa Lys	ata Ile	aag Lys	ata Ile 15	ggg	48
<22: <22: <400 cct Pro 1	2> (2 3> Po 0> 97 caa Gln caa	298) ortic 7 atc Ile ata	on of	HIV ctt t Leu 5 5	gg (Irp (caa o Gln <i>i</i>	cga Arg	ccc Pro	ctc Leu 10	gtc Val	Lys	Ile	Lys	Ile 15	Gly	4 8 96
<22: <22: <400 cct Pro 1 ggg Gly	2> (2 3> Po 0> 97 caa Gln caa Gln	298). Ortic 7 atc Ile ata Ile	act of Thr i	HIV ctt t Leu 5 5 gaa g	egg (Erp (gey t	caa o Sln i cta t Leu I	cga Arg ta (ccc Pro gat Asp 25	ctc Leu 10 aca Thr	gtc Val gga Gly	gca Ala	Ile gat Asp	Lys gat Asp 30	Ile 15 aca Thr	gtg Val	
<22: <22: <400 cct Pro 1 ggg Gly tta Leu	2> (2 3> Po 0> 97 caa Gln caa Gln gaa Glu	298). ortic atc Ile ata Ile gaa Glu 35	act of Thr :	HIV ctt t Leu 5 gaa c Glu 5 aat t Asn I	gg (frp (frp (frp (frp (frp (frp (frp (frp	caa o	ega Arg ta g Leu d gga a Sly 1	ccc Pro gat Asp 25 aaa Lys	ctc Leu 10 aca Thr tgg	gtc Val gga Gly aaa Lys	gca Ala cca Pro	gat Asp aaa Lys 45	gat Asp 30 ttg Leu	Ile 15 aca Thr ata Ile	gtg Val ggg Gly	96
<22: <22: <400 cct Pro 1 ggg Gly tta Leu gga Gly	2> (2 3> Po 0> 97 caa Gln caa Gln gaa Glu att Ile 50 atc	298). ortic atc Ile ata Ile gaa Glu 35 gga GGly	act of Thr :	HIV ctt t Leu 5 gaa c Glu 5 aat t ctt a he I	gcy taa I	caa caa caa caa caa caa caa caa caa caa	cga Arg tta g Leu ;	gat Asp 25 aaa Lys	ctc Leu 10 aca Thr tgg Trp	gtc Val gga Gly aaa Lys tat Tyr	gca Ala cca Pro gat Asp	gat Asp aaa Lys 45 Cag Gln	gat Asp 30 ttg Leu ata Ile	aca Thr ata Ile	gtg Val ggg Gly ata Ile	96 144

-177-

				Ile										Leu	aag Lys	3	336
		atg Met 115						Lys					Thr		gaa Glu	3	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	cta Leu	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ctg Leu 140	gaa Glu	aag Lys	gaa Glu	Gly 999	4	132
	Ile	tca Ser														4	80
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	5	28
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gag Glu	gtt Val	caa Gln 190	cta Leu	gga Gly	5	76
		cat His 195														6	24
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	tat Tyr 220	gag Glu	gac Asp	tty Phe	agg Arg	6	72
		act Thr														7	20
att Ile	agg Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	7	68
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	8	16
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	trt Xaa 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	8	64
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	9:	12
		cat His														96	60
cag Gln	aaa Lys	gaa Glu	Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	100	80

-178-

aa Ly	a tg s Tr	g acap Th	a gta r Vai 340	l Glı	g cci	t ata o Ila	a gta e Val	a cto l Leu 345) Pro	a gaa o Glu	a aaa a Lys	a gad s Asp	s ago Ser 350	Tr	g act o Thr	1056
gt Va	c aa l As:	t gad n Asp 355	p Ile	a cag e Glr	g aag 1 Lys	g tta Lev	gto Val 360	l Gly	a aaa / Lys	a ttg s Lev	aat Asr	tgg Trp 365	Ala	agt Ser	cag Gln	1104
		r Pro	a ggg													1116
<2:	10> ! 11> : 12> I 13> I	l115 DNA	ı Imm	unod	lific	ienc	y Vi	rus	(HIV	')						
<22	21 > C 22 > C	(0)	.(29 rote													
<22		298)	(on o			vers	e Tr	ansc	ript	ase						
cct	0> 9 caa Gln	atc	act Thr	cțt Leu 5	tgg T r p	caa Gln	cga Arg	ccc Pro	gtc Val 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	ggg Gly	48
999 999	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
tta Leu	gaa Glu	gaa Glu 35	atg Met	cat His	ttg Leu	cca Pro	gga Gly 40	aaa Lys	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly aaa	144
gga Gly	att Ile 50	Gly	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	cct Pro	gta Val	192
gaa Glu 65	aty Xaa	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct. Pro	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	Gly aaa	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384

-179-

aaa Lys	ata Ile 130	: Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	atg Met	Glu	aag Lys	gaa Glu	ggg Gly	432
	Ile					Pro					Asn				ttt Phe 160	480
					Āsp					Arg					ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	ttg Leu	gga Gly	576
		cat His 195														624
		ggt Gly														672
		act Thr														720
att Ile	aga Arg	tat Tyr	cag Gln	tat Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
		ttc Phe														816
		cca Pro 275														864
		gac Asp														912
aga Arg 305	caa Gln	cac His	ctg Leu	Leu	Lys	\mathtt{Trp}	Xaa	ttt Phe	Thr	Xaa	Pro	Asp	Lys	aaa Lys	cat His 320	960
cag Gln	aag Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gta Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gat Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104

-180-

		c tca r Se:														1	115
<2: ~<2:	10> 9 11> 3 12> I 13> I	L115 ONA	ı Imn	nunoc	lific	ienc	y Vi	.rus	(HIV	7)							
<22	21> C 22> C	CDS (0)															
<22		(298)			i) V Re	vers	e Tr	ansc	ript	ase							
cct	0> 9 cag	ato	act Thr	ctt Leu 5	tgg Trp	cag Gln	cga Arg	ccc	ctc Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	Gly		48
Gly 999	caa Gln	cta Leu	aag Lys 20	Glu	gct Ala	yta Xaa	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val		96
tta Leu	gaa Glu	gaa Glu 35	Met	aat Asn	ttg Leu	cca Pro	gga Gly 40	agr Xaa	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	GJY aaa	1	L44
gga Gly	att Ile 50	Gly	ggc Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	cag Gln	ata Ile	ccc Pro	cta Leu	1	192
gaa Glu 65	atc Ile	tgt Cys	ggc Gly	cat His	aag Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	2	40
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	act Thr	2	88
tta Leu	aat Asn	ttt Phe	ccc Pro 100	Ile	agt Ser	cct Pro	Ile	Glu	Thr	gta Val	cct Pro	Val	aaa Lys 110	Leu	aag Lys	3	36
cca Pro	gga Gly	atg Met 115	gat Asp	ggt Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	3	84
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gag Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	Gly aaa	4	32
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	4	80

-181-

gc Al	c at a Il	a aa e Ly	g aa 's Ly	a aaa s Lys 165	s Asp	agt Ser	act Thi	aaa Lys	tgg Trp	o Arg	a aa g Ly	a tta s Lei	a gta 1 Val	a gat l Asp 17!	t ttc p Phe	528
ag Ar	a ga g Gl	a ct u Le	t aa u As: 18	и гла	g aga Arg	act Thr	caa Glr	a gad Asp 185	Phe	tgg Trp	g ga o Gl	a gtt u Val	Caa Glr 190	ı Leı	a gga ı Gly	576
at: Il	a cc	a ca D Hi 19	s Pro	c tca Ser	Gly Gly	tta Leu	raa Xaa 200	Lys	aag Lys	, aaa : Lys	tca Sea	a gta val 205	Thr	gta Val	ctg Leu	624
ga! Asp	gtg Va. 210	r GT	t gat y Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asr 220	cca Pro	gat Asp	t.t.c Phe	agg Arg	672
aag Lys 225	туг	act Thi	t gca r Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	agg Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	tto Phe	caa Gln 260	agc Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	GIU	ata Ile	gtt Val	atc Ile	tac Tyr 280	caa Gln	tac Tyr	dtg Xaa	gat Asp	gat Asp 285	ttg Leu	tak Xaa	gta Val	864
rgc Xaa	tct Ser 290	gac Asp	tta Leu	gaa Glu	TTE	999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aag Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu '	tgg Trp	Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gtt Val 340	cag Gln	cct a Pro :	ata g Ile Y	vaı .	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	MSII	gac Asp 355	ata Ile	cag a Gln 1	aag 1 Lys 1	Leu I	gtg g Val (360	gga a Gly 1	aaa Lys :	ttg a Leu a	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile	tat Tyr 370	gca Ala	9 9	•												1115

<210> 100 <211> 1115

-182-

	12> 13>		ın Im	muno	difi	cien	cy V	irus	(HI	V)						
<2 <2		(0).	(2 Prot				-									
<2	21> : 22> 23> :	(298) ion ((111: of H:	5) IV R	evers	se Ti	ranso	crip	tase						
cct	00> : caa c Gli	a at	c act	t ctt r Lei 5	t tgg ı Trj	g caa o Glr	a cga n Arg	e ccc	cta Lei	ı Val	e aca	a ata r Ile	a aag	g ata s Ile 19	a gga e Gly	48
G1 ⁷	g cag / Glr	g cti n Xaa	r aag a Lys 20	GIU	a gct 1 Ala	ata lle	tta Lev	gat Asp 25	Thr	gga Gly	gca Ala	a gat a Asp	gat Asp 30	Thi	kta Xaa	96
tta Leu	gaa Glu	gaa Glu 35	ı Met	aat Asn	tng Xaa	r ccc	gga Gly 40	Arg	tgg Trp	ama Xaa	Pro	ama Xaa 45	Let	g ata i Ile	gly ggg	144
gga Gly	att Ile 50	: GTA	ggt Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	Gln	ata Ile	Pro	ata Ile	192
gaa Glu 65	тте	tgt Cys	gga Gly	cat His	aaa Lys 70	gtt Val	ata Ile	ggt Gly	aca Thr	gta Val 75	ttg Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct Pro	acc Thr	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	atg Met 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	Gly 999	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576

-183-

ata Ile	a cca Pro	cat His	Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	ata Ile	ctg Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aaa Lys 225	Val	tac Tyr	tgc Cys	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	acc Thr 235	aat Asn	gag Glu	acm Xaa	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctg Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	ccy Xaa	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	tgg Trp	agg Arg 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aag Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	Thr	gta Val 340	cag Gln	cct Pro	ata Ile	arg Xaa	ttg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	Asn	gam Xaa 355	ata Ile	cag Gln	aaa Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gcc Ala	agt Ser	cag Gln	1104
	tck Xaa 370	cng Xaa	3 3													1115
<211 <212	> 10 > 10 > DN > Hu	96	Immu	nodi	fici	ency	Vir	us (HIV)							
	> CD: > (0	·														

-184-

<2	21> 22> 23>	(298	3) :ion	.(109 of H	96) HIV R	.ever	se T	rans	crip	tase						
cc	00> t ca o Gl	r at	c ac e Th	t ct ir Le	u Tr	g ca p Gl:	g aco	c cc r Pro	c cti D Lei	ı Val	c yc. l Xa	a ata	a ag e Ar	g ak g Xa 1	a ggg a Gly 5	48
gg Xa	r ca a Gl	g yt n Xa	ацу	g ga s Gl	a gc u Al	t tta a Lei	a tta ı Lev	a gay ı Ası 29	Thi	a gra	a gca a Ala	a gat a Asp	t gar Ası 30	o Xa	a gta a Val	96
tt: Le:	a ga u Gli	a g <u>a</u> u Gl 3	u me	g ta t Ty	t tte	g cca	gga Gly 40	Arc	tgg Trp	, aas Lys	cca Pro	a aaa D Lys 45	Met	g ata	a ggg e Gly	144
gg;	a ato 7 Ile 50	5 GT.	a gg y Gl	t tti y Phe	t ato	aag Lys 55	Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	Glr	g ata	e Pro	ata Ile	192
gaa Glu 65	TTE	tgt	gg Gl	a cad y His	c aaa s Lys 70	: Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	ttg Leu	gta Val	gga Gly	tct Ser	aca Thr 80	240
Pro	gtt Val	aac Asr	ata n Ile	a att e Ile 89	GIY	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	att Ile	ggt Gly	tgc Cys 95	acc	288
tta Leu	aat Asn	ttt Phe	Pro) TTE	agt Ser	tct Ser	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aga Arg 110	tta Leu	aag Lys	336
ccc Pro	gga Gly	atg Met 115	Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	aag Lys	caa Gln	tgg Trp	cca Pro	tta Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gly ggg	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	ьys	aag Lys 165	Asn	agt Ser	gat Asp	aga Arg	tgg Trp 170	aga Arg	aaa Lys	gta Val	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	acc Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata [le	cca Pro	cat His 195	ccc Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	agg Arg	aga Arg	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
ъp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tac Tyr	ttt Phe 215	tca Ser	att Ile	ccc Pro	Leu .	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	aga Arg	672

-185-

aag Lys 225	: Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	Asr	gag Glu	g aca	cca Pro	ggg Gly 240	720
ato Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	Cag Gln 250	Gly	tgg Trp	, aas Lys	gga Gly	tca Sex 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	Arg	gaa Glu	816
cag Gln	aat Asn	cca Pro 275	gac Asp	atg Met	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	tta Leu	ttc Phe	aca Thr 315	cca Pro	gac Asp	caa Gln	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	ccg Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	act Thr	ata Ile	gtg Val	ctg Leu 345	cca Pro	gag Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gta Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	g			1096
<211 <212 <213)> 10 .> 10 !> DN !> Hu	48 A	Immu	nodi	fici	ency	Vir	us (HIV)							
<222	> CD > (0 > HI)														
<222	> CD: > (2: > Po:	98).	(10 n of	048) HIV	Rev	erse	Tra	nscr	ipta	se						
<400 cct Pro	cag a	atc a	act o	ctt i Leu '	tgg (Irp (cag (Gln 2	cga Arg	ccc Pro	tty Phe 10	gtc Val	aca Thr	ata Ile	aag Lys	gta Val 15	Gly 999	48
Gly (caa d Gln I	cta a Leu I	aag g Lys C 20	gaa g Slu <i>l</i>	gct (Ala 1	cta t Leu I	tg (Leu)	gat a Asp '	aca (gga (gca Ala	gat Asp	gat Asp 30	aca Thr	ata Ile	96

-186-

tta gaa gaa atg tgt ttg cca gga aga tgg aaa cca aaa ttg ata ggg 144 Leu Glu Glu Met Cys Leu Pro Gly Arg Trp Lys Pro Lys Leu Ile Gly gga att gga ggt ttt gtc aaa gta aga caa tat gat cag ata ccc ata 192 Gly Ile Gly Gly Phe Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile gaa atc tgt gga cat aaa gtt ata ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Val Ile Gly Thr Val Leu Val Gly Pro Thr 240 cct gcc aac ata gtt gga aga aat ctg ttg act cag att ggc tgt act Pro Ala Asn Ile Val Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 288 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag 336 Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 cca gga atg gat ggg cca aaa gtt aaa caa tgg cca ttg aca gaa gaa 384 Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu aaa ata aaa gca tta gta gaa att tgt aca gaa ttg gag aag gat gga 432 Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Leu Glu Lys Asp Gly aaa att tca aaa att ggg cct gaa aat cca tay aat act cca gta ttt 480 Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 528 Ala Ile Lys Lys Lys Asn Ser Asp Lys Trp Arg Lys Val Val Asp Phe 165 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtc caa tta gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 576 ata cca cat ccc gga ggg tta rag aag aac aaa tca ata aca gta ctg 624 Ile Pro His Pro Gly Gly Leu Xaa Lys Asn Lys Ser Ile Thr Val Leu gat gtg ggt gat gca tat ttt tca att ccc tta gat aaa gac ttc aga 672 Asp Val Gly Asp Ala Tyr Phe Ser Ile Pro Leu Asp Lys Asp Phe Arg 215 aag tat act gca ttt acc ata ccy agt ata aac aat gag aca cca ggg 720 Lys Tyr Thr Ala Phe Thr Ile Xaa Ser Ile Asn Asn Glu Thr Pro Gly att aga tat cag tat aat gtg ctt cca cag gga tgg aag gga tca cca 768 Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro 245 gcc ata ttc caa agt agc atg aca aaa ata tta gag cct ttt aga aag Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys

-187-

	a aat 1 Asr	Pro 27	o Asp	ata Ile	att Ile	atc Ile	gtt Val 280	caa Glr	tac Tyr	gtg Val	gat Asp	gat Asp 285	Lei	g ta u Ty:	t gta r Val	864
gca Ala	tct Ser 290	Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	Ile	aag Lys	g gaa s Glu	a cta 1 Leu	912
aga Arg 305	GIn	tat Tyr	ctg Leu	tgg Trp	gag Glu 310	tgg Trp	gga Gly	ttt Phe	tac Tyr	aca Thr 315	cca Pro	gac Asp	aaa Lys	a aaa s Lys	cat His 320	960
caa Gln	cag Gln	gaa Glu	ccc Pro	cca Pro 325	ttc Phe	ctc Leu	tgg Trp	atg Met	Gly 330	tat Tyr	gag Glu	ctc Leu	cat His	cct Pro	Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	a			1048
<211 <212	0> 10 L> 11 2> DI 3> Hi	116 NA	Immu	ınodi	fici	lency	Vir	as ((HIV)						,	
<220 <221 <222)> .> CI !> ((os))	. (297 rotea	')								-				
	> CI	s														
<223	> (2 > Po	98) ortic	on of	116) HIV	Rev	erse	Tra	nscr	ipta	se						
<223 <400 cct	> Po > 10 cag	rtic 3 atc	on of act	HIV	tgg	caa (cqa (ccc	ctc	atc	aca Thr	ata Ile	arg Xaa	rta Xaa 15	gly aaa	48
<223 <400 cct Pro 1	> Po > 10 cag Gln cag	ortic 3 atc Ile cta	on of	HIV ctt Leu ' 5	tgg Trp	caa (Gln)	cga Arg	ccc Pro	ctc Leu 10	gtc Val	Thr	Ile	Xaa	Xaa 15	Gly	48 96
<223 <400 cct Pro 1	> Pocag Gln cag Gln	ortico 3 atc Ile cta Leu	act Thr	HIV ctt Leu 5 gaa Glu aat	tgg Trp gct Ala	caa o Gln i cta t Leu i	cga Arg	ccc Pro gat Asp	ctc Leu 10 aca Thr	gtc Val gga Gly	gca Ala cca	Ile gat Asp aaa Lys	yat Asp 30	Xaa 15 aca Thr	gta Val	
<223 <400 cct Pro 1 ggg Gly tta	> Pocag Gln Cag Gln Gaa Glu Gaa	ortic 3 atc Ile cta Leu gaa Glu 35	act Thr aag Lys 20 atg	HIV ctt Leu 5 gaa g Glu aat t Asn I	tgg Trp gct Ala :	caa (Gln)	cga Arg	ccc Pro gat Asp 25 aga	ctc Leu 10 aca Thr	gtc Val	Thr gca Ala cca Pro	Ile gat Asp aaa Lys 45	yat Asp 30 atg Met	Xaa 15 aca Thr ata Ile	gta Val ggg Gly	96
<223 <400 cct Pro 1 ggg Gly tta Leu gga a	> Po > 10 cag Gln cag Gln gaa Glu att (15 11e 0	orticological state of the stat	act Thr aag Lys 20 atg Met Gly 1	HIV ctt Leu 5 gaa glu aat (Asn I	tgg Trp gct Ala: ttg Leu ittg	caa o	cga Arg	gat Asp 25 aga Arg	ctc Leu 10 aca Thr tgg a	gtc Val	gca Ala Cca Pro Sat Asp (gat Asp aaa Lys 45 cag	gat Asp 30 atg Met ata Ile	Xaa 15 aca Thr ata Ile	gta Val ggg Gly ata Ile	96 144

-188-

tt: Le:	a aat u Ası	ttt n Phe	ccc Pro	Ile	agt Ser	cct	att Ile	gaa Glu 105	ı Thi	gta Val	a cca Pro	gta Val	a aaa Lys	Lev	aag Lys		336
Pro	a gga o Gly	ate Met	: Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	. Lys	caa Glr	tgg Trp	p cca	cto Leu 125	Thr	gaa Glu	gaa Glu		384
aaa Lys	ata 130	Lys	gca Ala	tta Leu	aba Xaa	gaa Glu 135	Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	Glu	aag Lys	gaa Glu	ggr Xaa		432
aaa Lys 145	att Ile	tca Ser	aaa Lys	att Ile	ggg Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asn	act Thr	ccg Pro	gta Val	ttt Phe 160		4-8-0
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe		528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aaa Lys	act Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly		576
ata Ile	cca Pro	cac His 195	ccc Pro	gca Ala	Gly ggg	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu		624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	aaa Lys	gaa Glu	ttc Phe	agg Arg		672
aag Lys 225	tat Tyr	aca Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	aca Thr	ccc Pro	agg Arg 240		720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tcg Ser 255	cca Pro		768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys		816
caa Gln	aat Asn	cca Pro 275	Asp	Ile	gtt Val	Ile	Tyr	Gln	tat Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val		864
gga Gly	tct Ser 290	gac Asp	tta Leu	gag Glu	Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu		912
aga Arg 305	saa Xaa	cat His	ctg Leu	Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320		960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1	800

-189-

aaa Lys	tgg Trp	aca Thr	gtr Xaa 340	Glr	g cc	t ata	a rag	g ct a Le 34:	u Pr	a gaa o Gli	a aa ı Ly	a ga s As _l	c ag p Se 35	r Tr	g act p Thr	1056
gtc Val	Asn .	gac Asp 355	ata Ile	cag Gln	aaa Lys	a tta s Lei	a gtg 1 Va. 360	l Gl	a aaa y Lys	a tta 5 Lei	a aa 1 As:	t tgg n Trn 369	o Al	a ag a Se	t cag r Gln	1104
Ile	tac Tyr 370															1116
<210 <211 <212 <213	> DN2	16 4	Imm	unod	ific	ienc	y Vi	.rus	(HIV	7)						
<220: <221: <222: <223:	> CDS > (0)															
<2213 <2223 <2233	> (29	8).	(] n of	L116) E HIV) / Re	vers	e Tr	ansc	ript	ase						
<400 > cct c Pro G	cag a	tc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	mty Xaa 10	gtc Val	aca Thr	ata Ile	aag Lys	gta Val	Gly	48
ggg c	caa t Sln L	ta a eu 1	aaa Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
cta g Leu G	ilu G	aa a lu 1 35	ata Ile	aat Asn	ttg Leu	cca Pro	gga Gly 40	aga Arg	tgg Trp	aaa Lys	cca Pro	aaa Lys 45	atg Met	ata Ile	Gly aaa	144
gga a Gly I	itt g le G 50	ga g ly G	ggt Bly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	car Gln	ata Ile	cyt Xaa	ata Ile	192
gaa a Glu I 65	tc to	gt g ys G	ga Sly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct Pro	aca Thr 80	240
cct gr Pro Va	tc aa al As	ac a sn I	ita : le :	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttr Xaa 90	act Thr	cag Gln	att Ile	ggc Gly	tgc Cys 95	act Thr	288
tta aa Leu As	at ti sn Pi	ne P	ro :	ata : Ile :	agt Ser	cct Pro	Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca gg Pro Gl	ga at ly Me 11	et A	at g sp (ggc (Gly 1	cca Pro	Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384

-190-

aaa Lys	a ata s Ile 130	: Lys	a gca s Ala	tta Lev	a gya a Xaa	gaa Glu 135	ı Ile	tgt Cys	aca Thr	gaa Glu	ato Met	: Glu	aag Lys	gaa Glu	a gga ı Gly	432
aaa Lys 145	: Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	Asr	act Thr	cca Pro	gta Val	ttt Phe 160	480
gct Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gac Asp 185	Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cca Pro	gca Ala	gly aaa	cta Leu	cca Pro 200	agg Arg	aaa Lys	aga Arg	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	cca Pro	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	ccg Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gta Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gcc Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	att Ile	tta Leu	gat Asp	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	att Ile	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gca Ala	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cac His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gaa Glu	gaa Glu	cta Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	Phe	tac Tyr	Thr	Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttt Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	Thr	gta Val 340	cag Gln	cct Pro	ata Ile	Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104

-191-

att tat gca ggg Ile Tyr Ala Gly - 370	1116
<210> 105 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 105 cct cag atc act ctt tgg caa cga ccc ttc gtc gtc gta aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Phe Val Val Lys Ile Gly 1 5 10 15</pre>	48
ggg caa cta aag gaa gct cta tta gat aca gga gca gat aat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asn Thr Val 20 25 30	96
ttt gaa gac ytg aat ttg cca gga aaa tgg aaa cca aaa atg ata ggg Phe Glu Asp Xaa Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga cag tat gat cag gta ctt gta Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Val Leu Val 50 55 60	192
gaa atc tgt gga caa aaa gct ata ggt aca gta tta ata gga cct aca Glu Ile Cys Gly Gln Lys Ala Ile Gly Thr Val Leu Ile Gly Pro Thr 65 70 75 80	240
cct gtc aac ata att gga agg gat ctg ttg act cag att ggt tgc act Pro Val Asn Ile Ile Gly Arg Asp Leu Leu Thr Gln Ile Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135 140	432
aar att tca aaa att ggg cct gaa aac cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 145 150 155 160	480

-192-

gc Al	c ata a Ila	a aa e Ly	g aaa s Lys	a aas S Lys 165	Asp	agt Ser	act Thi	aar Lys	tgg Tri	o Arc	a aaa g Ly:	a tta s Lei	a gta 1 Va.	a gat l Asp 175	ttc Phe	528
ag Ar	a gaa g Gli	a ct: 1 Le:	t aat u Asr 180	і Гуз	g aga : Arg	act Thr	caa Glr	a gad a Asp 185	Phe	tgg Trp	gaa Glu	a gtt 1 Val	caa Glr 190	ı Let	a gga a Gly	576
ata Il	a cca e Pro	a cat His	s Pro	gca Ala	ggg Gly	tta Leu	aaa Lys 200	Lys	aaa Lys	a aaa s Lys	tca Ser	gta Val 205	Thr	gta Val	ctg Leu	624
ga! As <u>l</u>	ytg Val 210	. GT	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	Pro	tta Leu	gat Asp 220	Glu	gay Asp	ttc Phe	agg Arg	672
aag Lys 225	TAT	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agc Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	gga Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	tgt Cys	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gat Asp	cct Pro	ttt Phe 270	aga Arg	aag Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	cta Leu	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	rtg Xaa	gat Asp	gac Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gat Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	car Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aar Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	PIQ	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg T rp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag : Lys :	rea ,	gtg Val 360	gga Gly	aaa Lys	ttg Leu	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile	tac Tyr 370	cca Pro	ggg Gly													1116

<210> 106 <211> 1116

-193-

	?12> ?13>		n Im	muno	difi	cien	cy V	irus	(HI	V)						
<2 <2	20> 21> 22> 23>	(0).	(2 Prot	97) ease												
<2	21> 22> 23>	(298) ion	(111 of H	6) IV Re	ever	se Ti	cans	crip	tase						
CC.	00> t cas o Gli	gat	c ac e Th	t cti r Lei 5	t ngg ı Xaa	g caa a Glr	a ega n Arg	a cci g Xaa	n ati a Ile 10	e Val	c aca	a ata r Ila	a aag e Lys	g gta s Val	a ggg l Gly	48
999 Gly	g car Y Xaa	n tta a Lei	a aaa u Lys 20	S GIV	a gtt 1 Val	ytt Xaa	tta Lev	gat Asp 25) Xaa	a gga a Gly	gca Ala	a gat a Asp	gat Asp 30	Xaa	a gta a Val	96
tta Lei	a gaa 1 Glu	a gaa a Glu 35	ı Aac	gat Asp	ttg Leu	cca Pro	gga Gly 40	Arg	tgg Trp	aaa Lys	cca Pro	a aaa Lys 45	Met	ata Ile	Gly ggg	144
gga Gly	att Ile 50	: GTJ	a ggt / Gly	ttt Phe	atc Ile	aaa Lys 55	gta Val	aga Arg	cag Gln	tat Tyr	gat Asp 60	Gln	ata Ile	gtt Val	gta Val	192
gaa Glu 65	TTE	tgt Cys	gga Gly	cat His	aaa Lys 70	gct Ala	ata Ile	ggt Gly	aca Thr	gta Val 75	tta Leu	gta Val	gga Gly	cct	aca Thr 80	240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgc Cys 95	act Thr	288
	ASII	Pne	100	iie	Ser	Pro	lle	Glu 105	Thr	Val	Pro	gta Val	Lys 110	Leu	Lys	336
PIO	GIÀ	115	Asp	GIĀ	Pro	Lys	Val 120	Lys	Gln	Trp	Pro	ttg Leu 125	Thr	Glu	Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	ttg Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	att Ile	tca Ser	aaa Lys	aty Xaa	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
agg Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	Gin	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576

-194-

ata Ile	cca Pro	cat His	PIC	gc.	a gg a Gl	g yt y Xa	a aa a Ly: 20	e ra	g aa s As:	c aa n Ly	a tc s Se	a gt r Va 20	l Th	a gt r Va	a ctg l Leu	624
	210	Cly	voř	AIC	a ly	215	s sei	r val	l Pro	Lei	22	р L y: 0	s As _i	p Ph	t agg e Arg	672
225	- 3		1314	2 110	230	, 116	PIC) Ser	. 116	235	ı Ası	n Glu	ı Thi	r Pro	ggg Gly 240	720
-4- •	-5	-7-		245	No.	. vai	Leu	Pro	250	GIY	Tr	Lys	Gly	255		768
gca a Ala I			260		DCI	Mec	1111	265	TIE	Leu	GIU	Pro	270	Arc	,Lys	816
caa a Gln A		275			V41	116	280	GIII	ıyr	Met	Asp	285	Leu	Tyr	Val	864
gga t Gly S 2	ct c er 2 90	ac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gaa Glu	gaa Glu	ctg Leu	912
aga ge Arg Al 305	-u .		Deu	Deu	310	iip	GIÀ	Pne	Thr	Thr 315	Pro	Asp	Lys	Lys	His 320	960
cag aa Gln Ly				325	1116	Leu	пр	Met	330	Tyr	GIu	Leu	His	Pro 335	Asp	1008
aaa to Lys Tr		3	340	J_11	110	11 6	гуѕ	345	PTO	Glu	Lys	Asp	Ser 350	Trp	Thr	1056
gtc aa Val As		at a sp I 55	ata d [le (eag Sln	aag Lys	Deu	gtg Val 360	gga Gly :	aaa Lys	ttg Leu	Asn	tgg Trp 365	gcc Ala	agt Ser	cag Gln	1104
att ta Ile Ty 37	r P	ca g ro G	ga ly													1116
<210> <211> <211> <212> 1<213> 1	1116 DNA		ന്നസ്ഥാ	odif	icie	ency	Vir	ıs (H	IIV)				-			
<220> <221> (<222> (<223> F	(0).	(2 Prot	297) teas	e												

-195-

<221> CDS <222> (298)...(1116) <223> Portion of HIV Reverse Transcriptase <400> 107 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 48 ggg caa cta aag gaa gct tta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 96 tta gaa gaa atg gaa ttg cca gga aga tgg aaa cca aaa atg ata ggg Leu Glu Glu Met Glu Leu Pro Gly Arg Trp Lys Pro Lys Met Ile Gly 144 gga att gga ggt ttt atc aaa gta agm cag tat gat cag ata ccc ata GIy Ile GIy GIy Phe Ile Lys Val Xaa Gln Tyr Asp Gln Ile Pro Ile 192 gaa att tgt gga cat aaa gct gtg ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Val Gly Thr Val Leu Val Gly Pro Thr 240 cet gte aac ata att gga aga aat etg ttg act aag att ggt tge act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Lys Ile Gly Cys Thr 288 tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 336 105 cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 384 aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa gga Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 432 aaa att tca aaa att gga cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 480 gcc ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe 528 aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 576 185 ata cca cat ccc gca ggg tta aaa mgg aaa aaa tca gta aca gta ctg Ile Pro His Pro Ala Gly Leu Lys Xaa Lys Lys Ser Val Thr Val Leu 624 200 gat gtg ggt gat gca tat ttt tca gtt ccc tta gat aaa gag ttc agg Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Lys Glu Phe Arg 672 215

-196-

aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca gga Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly 235 240	720
att aga tat cag tac aat gtg yyt cca cag gga tgg aaa gga tca cca Ile Arg Tyr Gln Tyr Asn Val Xaa Pro Gln Gly Trp Lys Gly Ser Pro 245 250 255	768
gca ata ttc caa agt agc atg aca aaa atc tta gag cct ttt aga aaa Ala Ile Phe Gln Ser Ser Met Thr Lys Ile Leu Glu Pro Phe Arg Lys 260 265 270	816
caa aat cca gaa ata gtt atc tat cag tac atg gat gat ttg tat gta Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val 275 280 285	864
gga tct gac tta gaa ata ggg cag cac aga aca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu 295 300	912
aga caa cat ctg ttg aag tgg gga ttt acc aca cca gac aaa aaa cat Arg Gln His Leu Leu Lys Trp Gly Phe Thr Thr Pro Asp Lys Lys His 305 310 315	960
cag aaa gaa ccc cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008
aaa tgg aca gta cag cct ata gtg cta cca gaa aaa gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056
gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gcg agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tay gca ggg Ile Tyr Ala Gly 370	1116
<210> 108 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	
<pre><400> 108 cct caa atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48

-197-

ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gtg Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg aat ttg cca ggg aaa tgg aag cca aaa atg ata ggg Leu Glu Glu Met Asn Leu Pro Gly Lys Trp Lys Pro Lys Met Ile Gly 35 40 45	144
gga att gga ggg ttt atc aaa gta agm crg tat gat cag ata ccc ata Gly Ile Gly Gly Phe Ile Lys Val Xaa Xaa Tyr Asp Gln Ile Pro Ile 50 55 60	192
gaa atc tgt gra cat aaa gct aya ggt aca gta tta ata ggm cct act Glu Ile Cys Xaa His Lys Ala Xaa Gly Thr Val Leu Ile Xaa Pro Thr 65 70 75 80	240
cct gtc aac ata att gga aga awt ctg atg act cag att ggg tgc act Pro Val Asn Ile Ile Gly Arg Xaa Leu Met Thr Gln Ile Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gaa act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtt aaa caa tgg cca ttg aca gaa gag Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta gta gaa att tgt aca gaa atg gaa aag gaa ggg Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly 130 135 140	432
aag att tca aaa att ggg cct gaa aat cca tac aat act cca gta ttt Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Val Phe 150 155 160	480
gct ata aag aaa aaa gac agt act aaa tgg aga aaa tta gta gat ttc Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe 165 170 175	528
aga gaa ctt aat aag aga act caa gac ttc tgg gaa gtt caa tta gga Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly 180 185 190	576
ata cca cat cct gca ggt tta aaa aag aaa aaa tca gta aca gta cta Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser Val Thr Val Leu 195 200 205	624
gat gtg ggg gat gca tat ttt tca gtt ccc tta gat gaa aac ttc agg Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Asn Phe Arg 210 215 220	672
aag tat act gca ttt acc ata cct agt ata aac aat gag aca cca ggg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn Glu Thr Pro Gly 230 235 240	720
att aga tat cag tac aat gta ctt cca cag gga tgg aaa gga tca cca Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro 245 250 255	768

-198-

gca Ala	ata Ile	ttc Phe	caa Gln 260	tgt Cys	agc Ser	atg Met	aca Thr	aaa Lys 265	Ile	tta Leu	gag Glu	cct Pro	ttc Phe 270	Arg	aag Lys	816
caa Gln	aat Asn	cca Pro 275	gaa Glu	atg Met	gtt Val	atc Ile	trc Xaa 280	Gln	tac Tyr	gtg Val	gat Asp	gay Asp 285	ttg Leu	tat Tyr	gta Val	864
ggt Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctr Xaa	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Le u	agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctm Xaa	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	Inr	gtg Val 340	cag Gln	cat His	ata Ile	gaa Glu	ctg Leu 345	cca Pro	gaa Glu	caa Gln	gag Glu	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	yta Xaa	aat Asn	tgg Trp 365	gca Ala	agy Xaa	cag Gln	1104
att Ile	tat Tyr 370	gca Ala	ggg Gly													1116
<211 <212)> 10 .> 11 .> DN .> Hu	16 A	Immu	nodi	fici	ency	Vir	us (HIV)							
<222	> > CD > (0 > HI)	(297) oteas	se												
<222	> CD: > (2: > Po:	98)	(11 n of	116) HIV	Reve	erse	Tra	nscr	ipta	se						
<400 cct Pro	> 109 caa a Gln 1	atc a	ct c	tt t eu 7 5	gg (caa o	ga (Arg)	ccc (atc q Ile 1	gtc a Val '	aca (Thr	gta a Val 1	aag (Lys :	ata (Ile (gag Glu	48
Gly (cag d Gln I	ta a eu L	ag g ys G 20	aa c lu <i>A</i>	gct y Ala X	ta t aa I	ta d Leu)	gat a Asp : 25	aca o	gga g Gly 1	gca g Ala i	gat a Asp 1	aat a Asn 3	aca (Thr	gta Val	96
ttg d	gam g Kaa G	aa a lu I 35	ta a le A	at t sn I	tg c eu P	ca g ro G	ga a Sly 2 40	aga t Arg 1	gg a	aa d Lys I	cca a Pro 1	aaa a Lys N 45	itg a Met 1	ata g [le (ely eaa	144

-199-

998 Gl ₃	a att / Ile 50	2 Xaa	a ggt a Gly	ttt Phe	ato Ile	aaa Lys 55	: Val	aan Xaa	n cag a Gln	tat Tyr	gat Asp	Xaa	ata Ile	a mcc e Xaa	ata Ile	192	!
gad Asp	\circ Ile	tgt Cys	gga Gly	cat His	aaa Lys 70	Val	ata Ile	ggt	aca Thr	ata Ile 75	Leu	gta Val	gga	cct Pro	aca Thr 80	240	I
cct	gto Val	aac Asn	ata Ile	att Ile 85	Gly	aga Arg	gat Asp	ctg Leu	ttg Leu 90	Thr	cag Gln	att Ile	ggc Gly	tgc Cys 95	act	288	
tta Leu	aat Asn	ttt Phe	ccc Pro 100	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336	
cca Pro	gga Gly	atg Met 115	gat Asp	ggc	cca Pro	aaa Lys	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gar Glu	gaa Glu	384	
aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432	
aag Lys 145	att Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aac Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480	
gcc Ala	ata Ile	aag Lys	aaa Lys	aag Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528	
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576	
ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	Gly 999	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624	
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	tty Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gmt Xaa 220	aaa Lys	gaa Glu	tnn Xaa	nnn Xaa	672	
nnn Xaa 225	nnn Xaa	nnn Xaa	nnn Xaa	Xaa	nnn Xaa 230	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa 235	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa 2 4 0	720	
nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa	nnn Xaa 245	nnn Xaa	nnn Xaa	nnn Xaa	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768	
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816	
caa Gln	aat Asn	cca Pro 275	gaa Glu	ata Ile	gtt Val	Ile	tac Tyr 280	car Gln	tac Tyr	rtg Xaa	gat Asp	gay Asp 285	ttg Leu	ttw Xaa	gta Val	864	

-200-

G1 99	a tci y Sei 290	r Asr	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	= Ile	gag Glu	ggaa Glu	ctg Leu	912
ag Ar ~30	g Glr	a cat n His	ctg Leu	ttg Leu	agg Arg 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
Ca Gl	g aaa n Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	99y Xaa 330	tat Tyr	gaa Glu	ctc Leu	cat His	ect Pro 335	gat Asp	1008
aa: Ly:	a tgg s Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	Trp	act Thr	1056
gto Val	e aat l Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile	tat Tyr 370	Pro	ggg													1116
<21 <21	.0> 1 .1> 1 .2> DI .3> H	116 NA	Immu	ibonı	fici	ency	, Vii	rus ((HIV)					-		
<22	1> Cl 2> (0)	. (297 rotea													
<22		298).	(1 on of			erse	Tra	nscr	·ipta	se						·
cyt	0> 11 cag Gln	atc	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	cts Xaa 10	gtc Val	aca Thr	ata Ile	aag Lys	gta Val 15	gly aaa	48
Gly aaa	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	Asp	aca Thr	gga Gly	gca Ala	gat Asp	Asp	aca Thr	gta Val	96
			20					25					30			
tta Leu	gaa Glu	gaa Glu 35	atr . Xaa .	aat Asn	ttg Leu	cca (Pro :	ggr Xaa 40	aaa	tgg Trp	aaa Lys	cca Pro	awa Xaa 45	atq	ata Ile	Gly aaa	144
Leu gga	Glu	Glu 35 gga	atr	Asn ttt	Leu acc	Pro :	Xaa 40 qta	aaa Lys aga	Trp cag	Lys tat (Pro gat	Xaa 45 cag	atg Met	Ile	Gly	144 192

-201-

cct Pro	gto Val	aac Asr	ata 1 Ile	att : Ile :85	: Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	Thr	cag Glr	g att	ggt Gly	tgc Cys 95		288
tta Leu	a aat a Asn	ttt Phe	ccc Pro 100	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336
cca Pro	ggg Gly	Met 115	Asp	ggc Gly	cca Pro	aaa Lys	gtt Val 120	Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	ata Ile	gaa Glu 135	atc Ile	tgt Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
					gac Asp											528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	ggg ggg	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gtg Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tac Tyr	act Thr	gca Ala	ttt Phe	mcc Xaa 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	Phe	caa Gln 260	agt Ser	agc Ser	Met	Thr	aaa Lys 265	Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	Arg	aaa Lys	816
caa Gln	mat Xaa	cca Pro 275	gac Asp	atg Met	gty Xaa	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	ggr Xaa 295	cag Gln	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	cag Gln	cat His	ttg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aag Lys	cat His 320	960

-202-

GI	g aa n Ly	a ga s Gl	g cct u Pro	cca Pro 325) Pne	ctt Lev	tgg Tr	g ato P Met	g gg(: Gl) 330	y Ty:	t gaa	a ct u Le	c car u His	t cc s Pr	o Asr	1008
aa Ly	a tg s Tr	g aca p Thi	a gta r Val 340	. Gin	cct Pro	ata Ile	gag Glu	g ctg Let 345	ı Pro	a gaa o Glu	a aas 1 Lys	r gai s Xaa	m ago a Sei 350	r Trj	g act o Thr	1056
gt Va	c aa l As	t gad n Ası 35!	c ata p Ile	cag Gln	aaa Lys	ata Ile	gto Val 360	. Gly	aaa Lys	tto Lev	aat Asr	tgg Trp 369	o Ala	a agt	cag Gln	1104
at:	t tad e Tyn 370	r Pro	g Gly													1116
<21 <21 <21		116 NA	ı Immı	unod:	ific	iency	y Vi:	rus	(HIV)						
<22	21> C 22> (0)	.(291 rotea													
<22	1> C 2> (3> P	298)	(I	1116) HIV	7 Rev	verse	e Tra	ansci	ripta	ase						
<40																
	0> 1		act	ctt	taa	C22	cas		a+ a							
cct Pro 1	cag Gln	atc	act Thr	Leu 5	Trp	Gln	Arg	Pro	Leu 10	Val	Thr	Ile	Lys	Ile 15	Gly	48
cct Pro 1 ggg	cag Gln caa	atc Ile ata	act Thr aag Lys 20	5 gaa	Trp qct	Gln cta	Arg	Pro	Leu 10	Val	Thr	Ile	Lys	Ile 15	Gly	48 96
cct Pro 1 999 Gly	cag Gln caa Gln	atc Ile ata Ile	aag Lys	gaa Glu agc	gct Ala	Cta Leu Cca	tta Leu	Pro gat Asp 25	Leu 10 aca Thr	Val gga Gly	Thr gca Ala	Ile gat Asp	gat Asp 30	Ile 15 aca Thr	Gly gta Val	
cct Pro 1 ggg Gly tta Leu	cag Gln caa Gln gaa Glu	atc Ile ata Ile gaa Glu 35	aag Lys 20	gaa Glu agc Ser	gct Ala ttg Leu	cta Leu cca Pro	tta Leu gga Gly 40	gat Asp 25 aaa Lys	Leu 10 aca Thr tgg Trp	yal gga Gly aaa Lys	Thr gca Ala cca Pro	gat Asp aaa Lys 45	gat Asp 30 atg Met	Ile 15 aca Thr ata Ile	gta Val ggg Gly	96
ggg Gly tta Leu gga Gly	cag Gln caa Gln gaa Glu att Ile 50	atc Ile ata Ile gaa Glu 35 gga Gly	aag Lys 20 atg Met	gaa Glu agc Ser ttt Phe	gct Ala ttg Leu atc	cta Leu cca Pro aaa Lys 55	tta Leu gga Gly 40 gta Val	gat Asp 25 aaa Lys agm Xaa	Leu 10 aca Thr tgg Trp cag Gln aca	yal gga Gly aaa Lys tat Tyr	Thr gca Ala cca Pro gwt Xaa 60	gat Asp aaa Lys 45 cat His	gat Asp 30 atg Met ata Ile	Ile 15 aca Thr ata Ile ccc Pro	gta Val ggg Gly ata Ile	96 144
ggg Gly tta Leu gga Gly gaa Glu 65	cag Gln caa Gln gaa Glu att Ile 50 wtc Xaa	atc Ile ata Ile gaa Glu 35 gga Gly tgt Cys	aag Lys 20 atg Met ggt Gly	gaa Glu agc Ser ttt Phe	gct Ala ttg Leu atc Ile aaa Lys	cta Leu cca Pro aaa Lys 55 gct	tta Leu gga Gly 40 gta Val	gat Asp 25 aaa Lys agm Xaa ggt Gly	Leu 10 aca Thr tgg Trp cag Gln aca Thr	yal gga Gly aaa Lys tat Tyr gta Val 75	Thr gca Ala cca Pro gwt Xaa 60 tta Leu	gat Asp aaa Lys 45 cat His ata Ile	gat Asp 30 atg Met ata Ile	Ile 15 aca Thr ata Ile ccc Pro	gta Val ggg Gly ata Ile aca Thr 80	96 144

-203-

Pro	a gga o Gly	a at y Me 11	t Ası	Gly Gly	g cca y Pro	a aaa o Lys	gti Val	l Lys	g caa s Gli	a tgg n Trj	g cc p Pr	a ct o Le 12	u Th	a ga r Gl	a gaa u Glu	384
aaa Lys	a ato 5 Ile 130	Ly:	a gca s Ala	a ttg a Lei	g ata 1 Ile	gaa Glu 135	ı Ile	tgt Cys	aca Thi	a gaa r Glu	a at 1 Mer 14	t Gl	a aag 1 Lys	g ga s Gl	a gga u Gly	432
aaa Lys 145	: TTE	gaa Glu	a aaa 1 Lys	att Ile	999 Gly 150	Pro	gaa Glu	a aat 1 Asn	cca Pro	tac Tyr 155	: Ası	t act	cca Pro	a gta	a ttt l Phe 160	480
gcc	ata Ile	agg Arg	g aaa g Lys	aaa Lys 165	Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	Arg	aaa J Lys	a tta s Lev	gta Val	gai Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Lev	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gaa Glu	a att	caa Gln 190	Let	gga Gly	576
ata Ile	cca Pro	cat His 195	Pro	gca Ala	Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	gta Val	aat Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	gga Gly 240	720
att Ile	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	caa Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gaa Glu	yta Xaa	gtt Val	Ile	tac Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tca Ser 290	gac Asp	tta Leu	Glu	ata Ile	Glu	Lys	cat His	aga Arg	gca Ala	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	gaa Glu	cat His	ctg Leu	хаа	aaa Lys 310	tgg Trp	gly ggg	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttt Phe	ctt Leu	tgg Trp	Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	acc : Thr :	ata a Ile i	Lys	ctg Leu : 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056

-204-

gtc aat gat ata cag aag tta gtg gga aaa ttg aat tgg gca agt caa Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104
att tat cca ggg Ile Tyr Pro Gly 370	1116
<210> 112 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)	
<220> <221> CDS <222> (0)(297) <223> HIV Protease	
<221> CDS <222> (298)(1116) <223> Portion of HIV Reverse Transcriptase	·
<pre><400> 112 cct cag atc act ctt tgg caa cga ccc ctc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Leu Val Thr Ile Lys Ile Gly 1 5 10 15</pre>	48
ggg cag cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96
tta gaa gaa atg aat ttg cca gga aga tgg aaa cca aaa atk ata ggg Leu Glu Glu Met Asn Leu Pro Gly Arg Trp Lys Pro Lys Xaa Ile Gly 35 40 45	144
gga att gga ggt ttt atc aaa gta aga cag tat gat cag ata ctt gta Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Leu Val 50 55 60	192
gaa att tgt gga cat aaa gct ata ggt aca gta tta gta gga cct aca Glu Ile Cys Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr 65 70 75 80	240
cct gtc aac ata att gga aga aat ctg ttg act cag att ggt tgc act Pro Val Asn Ile Ile Gly Arg Asn Leu Leu Thr Gln Ile Gly Cys Thr 85 90 95	288
tta aat ttt ccc att agt cct att gag act gta cca gta aaa tta aag Leu Asn Phe Pro Ile Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys 100 105 110	336
cca gga atg gat ggc cca aaa gtc aaa caa tgg cca ttg aca gaa gaa Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu 115 120 125	384
aaa ata aaa gca tta atg gaa att tgt gca gaa wtg gaa aag gaa gga Lys Ile Lys Ala Leu Met Glu Ile Cys Ala Glu Xaa Glu Lys Glu Gly 130 135 140	432

-205-

aa Ly 14	s Ile	t tca e Sei	a aaa c Lys	att Ile	999 Gly	Pro	gaa Glu	a aat 1 Asn	cca Pro	tac Tyr 155	Asr	act Thi	cca Pro	a gta o Val	a ttt l Phe 160	480
gc Al	c ata a Ile	a aag e Lys	g aaa E Lys	aaa Lys 165	Asp	ago Ser	act Thr	aaa Lys	tgg Trp 170	Xaa	aaa Lys	tta Lev	gta Val	gat Asp 175	ttc Phe	528
aga Arg	a gaa g Glu	ctt Leu	aat Asn 180	Lys	aga Arg	act Thr	caa Gln	gac Asp 185	Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	Lev	gga Gly	576
ata Ile	e Pro	cat His 195	Pro	gca Ala	Gly	tta Leu	aaa Lys 200	Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	Thr	gta Val	cta Leu	624
gat Asp	gtg Val 210	Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gat Asp 220	gaa Glu	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aac Asn 235	aat Asn	gag Glu	acm Xaa	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	aca Thr	aaa Lys 265	att Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	cag Gln	cat His	ctg Leu	ttg Leu	aag Lys 310	tgg Trp	gga Gly	ttk Xaa	tmc Xaa	aca Thr 315	cca Pro	gac Asp	aaa Lys	aaa Lys	cat His 320	960
cag Gln	aaa Lys	saa Xaa	Pro	cca Pro 325	ttc Phe	Leu	\mathtt{Trp}	atg Met	Gly	\mathtt{Tyr}	Glu	Leu	Xaa	cct Pro 335	Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	caa Gln	cct Pro	ata Ile	Val	ctg Leu 345	cca Pro	gaa Glu	aag Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys :	Leu '	gtg Val 360	gga Gly	aaa Lys	ttr Xaa	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att Ile	tac Tyr 370	gca Ala	ggg ggg													1116

-206-

<2: <2:	10> 1 11> 1 12> I 13> F	116 NA	ı Imn	unoć	lific	ienc	y Vi	.rus	(HIV	7)						
<22	21> C	0)	.(29 rote	7) ase												
<22	1> C 2> (3> P	298)	(on o	1116 f HI) V Re	vers	e Tr	ansc	ript	ase						
cct	0> 1 cag Gln	ato	act Thr	ctt Leu 5	tgg Trp	caa Gln	cga Arg	ccc Pro	ctc Leu 10	gtc Val	aca Thr	ata Ile	aag Lys	ata Ile 15	Gly ggg	48
ggg Gly	caa Gln	cta Leu	aag Lys 20	gaa Glu	gct Ala	cta Leu	tta Leu	gat Asp 25	aca Thr	gga Gly	gca Ala	gat Asp	gat Asp 30	aca Thr	gta Val	96
												aaa Lys 45				144
												cag Gln				192
												ata Ile				240
cct Pro	gtc Val	aac Asn	ata Ile	att Ile 85	gga Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	act Thr	cag Gln	ctt Leu	ggt Gly	tgt Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggt Gly	cca Pro	aga Arg	gtt Val 120	aaa Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	acm Xaa	gaa Glu	gaa Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	ata Ile	gaa Glu 135	atc Ile	tgc Cys	aca Thr	gaa Glu	atg Met 140	gaa Glu	aag Lys	gam Xaa	sga Xaa	432
waa Xaa 145	att Ile	tca Ser	aaa Lys	mta Xaa	999 Gly 150	cct Pro	gam Xaa	wat Xaa	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gta Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528

-207-

aga gaa ctt Arg Glu Leu			Gln								576
ata cca cac Ile Pro His	Pro Ala										624
gat gtg ggt Asp Val Gly 210			Ser								672
aag tat act Lys Tyr Thr 225	gca ttt Ala Phe	acc ata Thr Ile 230	cct a	agt at Ser Il	a aac e Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
atc aga tat Ile Arg Tyr	cag tac Gln Tyr 245	aat gtg Asn Val	ctt (Leu)	cca ca Pro Gl: 25	a Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca ata ttc Ala Ile Phe			Thr I								816
caa aat cca Gln Asn Pro 275	gaa ata Glu Ile	gtt atc Val Ile	tgt o Cys 0 280	caa ta Gln Ty:	e atg	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga tct gac Gly Ser Asp 290											912
aga gaa cat Arg Glu His 305	ctg ttg Leu Leu	aag tgg Lys Trp 310	gga t Gly F	tt acc	aca Thr 315	cca Pro	gat Asp	aaa Lys	aaa Lys	cat His 320	960
cag aaa gaa Gln Lys Glu	cct cca Pro Pro 325	ttc ctt Phe Leu	tgg a Trp M	tg ggt let Gly 330	Tyr	gag Glu	ctc Leu	cat His	cct Pro 335	gat Asp	1008
aaa tgg aca Lys Trp Thr	gta cag Val Gln 340	cct ata Pro Ile	Val L	tg cca eu Pro 45	gaa Glu	aaa Lys	Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc aat gac Val Asn Asp 355	ata cag Ile Gln	aag tta Lys Leu	gtg g Val G 360	ga aaa ly Lys	ttg Leu	Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
att tat gca Ile Tyr Ala 370	Gly 999										1116
<210> 114 <211> 1116 <212> DNA <213> Human	Immunodi	ficiency	Viru	s (HIV)						

<220>

-208-

<221> CDS <222> (0). <223> HIV					
<221> CDS <222> (298) <223> Port:)(1116) ion of HIV R	everse Trans	scriptase		
<400> 114 cmt caa atm	n amt ctt tg a Xaa Leu Tr 5	g car mra co p Gln Xaa Pr	cc cta gtc o co Leu Val : 10	cna awn nmm g Xaa Xaa Xaa X	kk agg 48 aa Arg 15
ggg gca aat Gly Ala Asr	aag gaa gc Lys Glu Ala 20	a Leu Leu As	ac aca gga g sp Thr Gly 2 55	gca gat gat m Ala Asp Asp X 30	ca gta 96 aa Val
tta gaa gaa Leu Glu Glu 35	ı met xaa Lei	a cca gga aa 1 Pro Gly Ly 40	a tgg aaa d s Trp Lys I	cca aaa atg a Pro Lys Met I 45	ta ggg 144 le Gly
gga att gga Gly Ile Gly 50	ggt ttt ato Gly Phe Ile	aaa gta ag Lys Val Xa 55	n cag tat g a Gln Tyr G	gag cag ata c Slu Gln Ile P 60	cc ata 192 ro Ile
gaa atc tgt Glu Ile Cys 65	gga cat aaa Gly His Lys 70	Ala Ile Gl	t aca gta t y Thr Val I 75	tg gta ggm c Leu Val Xaa P	ct aca 240 ro Thr 80
cct gtc aac Pro Val Asn	ata att gga Ile Ile Gly 85	aga aat cto Arg Asn Leo	g ttg act c u Leu Thr G 90	ag att ggt to In Ile Gly C	gc act 288 /s Thr 95
tta aat ttt Leu Asn Phe	ccc att agt Pro Ile Ser 100	cct att gas Pro Ile Glu 105	u Thr Val P	ca gtg aaa tt Pro Val Lys Le 110	a aag 336 u Lys
cca gga atg Pro Gly Met 115	gat ggc cca Asp Gly Pro	aaa gtt aaa Lys Val Lys 120	a caa tgg c s Gln Trp P	ca tta aca ga ro Leu Thr Gl 125	a gaa 384 u Glu
aaa ata aaa Lys Ile Lys 130	gca tta gta Ala Leu Val	gaa att tgt Glu Ile Cys 135	s Thr Glu M	tg gaa aaa ga et Glu Lys Gl 40	a ggg 432 u Gly
aaa att tca Lys Ile Ser 145	aaa att ggg Lys Ile Gly 150	cct gaa aat Pro Glu Asn	cca tac ac Pro Tyr A 155	at act cca gt sn Thr Pro Va	a ttt 480 l Phe 160
gcc ata aag Ala Ile Lys	aaa aaa gac Lys Lys Asp 165	agt act aaa Ser Thr Lys	tgg aga ag Trp Arg Ly 170	aa tta gta ga ys Leu Val As 17	p Phe
ard Gin ren	aat aag aga Asn Lys Arg 180	act caa gac Thr Gln Asp 185	Phe Trp G	aa gtc caa tt lu Val Gln Le 190 .	a gga 576 u Gly
ata cca cat Ile Pro His 195	cct gca ggg Pro Ala Gly	tta aaa aag Leu Lys Lys 200	aaa aaa to Lys Lys Se	ca gta aca gter er Val Thr Va 205	g ctg 624 l Leu

-209-

gac gtg ggt Asp Val Gly 210							672
aag tat act Lys Tyr Thr _225	Ala Phe X						720
agt agg tat Ser Arg Tyr							768
gca ata ttc Ala Ile Phe							816
caa aat cca Gln Asn Pro 275							864
gga tct gac Gly Ser Asp 290							912
aga caa cat Arg Gln His 305	Leu Leu A						960
cag aar gaa Gln Lys Glu							1008
aaa tgg aca Lys Trp Thr	gta cag co Val Gln Pr 340	ct ata gtg co Ile Val	ctg cca Leu Pro 345	gaa aaa Glu Lys	gac ags Asp Xaa 350	ttg rct Leu Xaa	1056
kca aat gac Xaa Asn Asp 355							1104
att tac tca (Ile Tyr Ser (370							1116
<210> 115 <211> 1116 <212> DNA <213> Human	Immunodifi	.ciency Vir	rus (HIV)				
<220> <221> CDS <222> (0) <223> HIV Pro							
<221> CDS <222> (298). <223> Portion		everse Tra	nscripta	se			

-210-

CC	OGI	g a	tc a le 1	ict Thr	ctt Leu 5	tg: Tr	g ca o Gli	a cg n Ar	ja co g Pr	O L	ະເວີດ ອີນ ໄ	gtc /al	ac. Th:	a at r Il	a aa e Ly	ag a 's]	ata [le	gly ggg	48
99 Gl	g ca y Gl	g ci n Le	ca a eu I	ag ys 20	gaa Glu	gct Ala	cta Lei	a at ı Il	e As	t ac p Th	ea g ur G	gga Sly	gca Ala	a ga a As	p As	t a p T	ca hr	gtg Val	96
tt Le	a ga u Gl	u G	aa a lu M 85	tg et	agt Ser	ata Ile	cca Pro	gg Gl:	у Lу	a to s Tr	jg a	aa ys	cca Pro	a aaa D Lys 45	s Le	g a u I	ta le	ggg Gly	144
gg: Gl	a at y Ilo 50	5 G1	jā g y G	gt ly	ttt Phe	atc	aaa Lys 55	Va.	a ag l Ar	a ca g Gl	g t n T	at yr	gat Asp 60	cag Glr	g gk	g c a P	cc ro	gta Val	192
gaa Glu 65	1 116	t tg e Cy	t g s G	ga ly	cat His	aaa Lys 70	gct Ala	ata Ile	a ggi e Gly	t mc y Xa	a X	tw aa 75	tta Leu	ata Ile	a ggi e Xaa	m c a P	ct ro	aca Thr 80	240
cct Pro	gco Ala	aa As	c at n II	le :	att Ile 85	gga Gly	agg Arg	aat Asr	cto Lev	g ttg 1 Lem 9	u Tl	ct hr	cag Gln	att Ile	ggt Gly	/ C	gc ys 95	act Thr	288
tta Leu	aat Asn	tt Ph	t co e Pr 10	. 0	att Ile	agt Ser	cct Pro	att	gaa Glu 105	Th	t gt r Va	al	cca Pro	gta Val	aaa Lys 110	Le	a	aag Lys	336
cca Pro	gga Gly	Mei	- AS	it g	ggc Sly	cca Pro	aaa Lys	gtt Val 120	Lys	caa Glr	a to	.b	cca Pro	ttg Leu 125	aca Thr	ga G]	aa .u	gag Glu	384
Буб	130	пλг	s AI	a I	Jeu	inr	135	IIe	Cys	Thr	Gl	.u !	Met 140	gaa Glu	Lys	G]	u (Gly	432
aag Lys 145	att Ile	tca Ser	aa Ly	a a s I	те (999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	ta Ty 15	r 1	aat Asn	act Thr	cca Pro	gt Va	1	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aa. Ly:	ון פ	aa q ys 1 65	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	ag Ar	a a g I	aaa Lys	tta Leu	gta Val	ga As 17	p I	tc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asi 180	ı	ag a ys A	aga Arg	act Thr	caa Gln	gat Asp 185	ttc Phe	tg Tr	p 0	gaa Slu	gtt Val	caa Gln 190	tt Le	a c u G	ga Sly	576
ata Ile	cca Pro	cat His 195	Pro	g A	ca c la c	igg :	Leu	aaa Lys 200	aag Lys	aaa Lys	aa: Ly:	a t	er	gta Val 205	aca Thr	gt. Va	a c	tg eu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	. g	ca t la T	yr 1	ttt Phe : 215	tca Ser	gtt Val	ccc Pro	tta Lei	ιA	sp 20	gaa Glu	gac Asp	tt: Phe	ta ≥A	19 99	672
aaa Lys 225	tat Tyr	act Thr	gca Ala	t t	ie i	cc a hr 1 30	ata ([le]	ect Pro	agt Ser	gta Val	aac Asr 235	1 A	at (gag Glu	aca Thr	cca Pro	G	99 ly 40	720

-211-

at:	aga Arg	tat Tyr	cag Gln	tat Tyr 245	Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255		768
gca Ala	ata Ile	ttc Phe	caa Gln 260	tgt Cys	agt Ser	atg Met	aca Thr	aaa Lys 265	ata Ile	tta Leu	gag Glu	ccc	ttt Phe 270	Arg	aaa Lys	816
ca <i>a</i> Glr	aat Asn	cca Pro 275	gac Asp	cta Leu	gtt Val	atc Ile	tat Tyr 280	caa Gln	tac Tyr	gtg Val	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	Ser 290	gac Asp	tta Leu	gaa Glu	ata Ile	999 Gly 295	cag Gln	cat His	aga Arg	aca Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912
aga Arg 305	caa Gln	cat His	ctg Leu	ttg Leu	aaa Lys 310	tgġ Trp	ggt Gly	ttt Phe	acc Thr	aca Thr 315	cca Pro	gac Asp	aaa Lys	aag Lys	cat His 320	960
cag Gln	aaa Lys	gaa Glu	cct Pro	cca Pro 325	ttc Phe	ctt Leu	tgg Trp	atg Met	ggt Gly 330	tat Tyr	gaa Glu	ctc Leu	cat His	cca Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	cag Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	cca Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	tgg Trp	act Thr	1056
gtc Val	aat Asn	gac Asp 355	ata Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	tta Leu	aat Asn	tgg Trp 365	gca Ala	agt Ser	cag Gln	1104
	tac Tyr 370															1116
<210> 116 <211> 1116 <212> DNA <213> Human Immunodificiency Virus (HIV)																
<220> <221> CDS <222> (0)(297) <223> HIV Protease																
<222	> CD: > (2: > Po:	98).			Rev	erse	Tra	nscr	ipta	se						
cct	> 110 cag a Gln :	atc a	act o Thr I	ett (Leu :	tgg (Trp (caa (Gln)	cga (Arg)	ccc (Pro 1	ctc Leu 10	gtc a Val '	aca Thr	ata Ile	aag Lys	ata Ile 15	gly ggg	48
Gly	cag o	cta a Leu I	ag g Lys C 20	gaa g Slu <i>l</i>	gct d Ala I	cta 1 Leu 1	tta q Leu <i>i</i>	gat a Asp : 25	aca q Thr (gga q Gly i	gca g Ala	gat Asp	gac Asp 30	aca Thr	gta Val	96

-212-

			ı Ile					Arg					Lev		ggg Gly	144
		Gly										Gln			ata Ile	192
	Ile					Ala					Leu				aca Thr 80	240
															act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro 100	att Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	act Thr	gta Val	cca Pro	gta Val	aaa Lys 110	tta Leu	aag Lys	336
cca Pro	gga Gly	atg Met 115	gat Asp	ggc Gly	cca Pro	aag Lys	gtt Val 120	aag Lys	caa Gln	tgg Trp	cca Pro	ttg Leu 125	aca Thr	gaa Glu	gaa [.] Glu	384
aaa Lys	ata Ile 130	aaa Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	Gly 999	432
						cct Pro										480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	aca Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttt Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	cta Leu	Gly ggg	576
ata Ile	cca Pro	cat His 195	ccc Pro	gca Ala	999 Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	ttg Leu	gat Asp 220	aaa Lys	gac Asp	ttc Phe	agg Arg	672
aag Lys 225	tac Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	ata Ile	aat Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	caa Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agt Ser	agc Ser	atg Met	Thr	aaa Lys 265	atc Ile	tta Leu	gag Glu	cct Pro	ttt Phe 270	aga Arg	aaa Lys	816

WO 01/35316 PCT/US00/30863

-213-

caa aat cca gac ata gtt atc tat caa tac gta gat gac ttg tat gta Gln Asn Pro Asp Ile Val Ile Tyr Gln Tyr Val Asp Asp Leu Tyr Val 275 280 285	864													
gga tct gac tta gaa ata ggg cag cat aga aca aaa ata gag gaa ctg Gly Ser Asp Leu Glu Ile Gly Gln His Arg Thr Lys Ile Glu Glu Leu 290 295 300	912													
aga caa cat ctg tgg aag tgg ggg ttt tac aca cca gat aaa aaa cat Arg Gln His Leu Trp Lys Trp Gly Phe Tyr Thr Pro Asp Lys Lys His 305 310 315 320	960													
cag aaa gaa cct cca ttc ctt tgg atg ggt tat gaa ctc cat cct gat Gln Lys Glu Pro Pro Phe Leu Trp Met Gly Tyr Glu Leu His Pro Asp 325 330 335	1008													
aaa tgg aca gta cag cct ata gtg ctg cca gaa aag gac agc tgg act Lys Trp Thr Val Gln Pro Ile Val Leu Pro Glu Lys Asp Ser Trp Thr 340 345 350	1056													
gtc aat gac ata cag aag tta gtg gga aaa tta aat tgg gca agt cag Val Asn Asp Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln 355 360 365	1104													
att tac cca ggg Ile Tyr Pro Gly 370	1116													
<210> 117 <211> 1119 <212> DNA <213> Human Immunodificiency Virus (HIV)														
<pre><220> <221> CDS <222> (0)(297) <223> HIV Protease</pre>														
<pre><221> CDS <222> (298)(1119) <223> Portion of HIV Reverse Transcriptase</pre>														
<400> 117 cct caa atc act ctt tgg caa cga ccc atc gtc aca ata aag ata ggg Pro Gln Ile Thr Leu Trp Gln Arg Pro Ile Val Thr Ile Lys Ile Gly 1 5 10 15	48													
ggg caa cta aag gaa gct cta tta gat aca gga gca gat gat aca gta Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val 20 25 30	96													
tta gaa gaa atg gat ttg cca gga aga tgg aca cca aaa atg ata ggg Leu Glu Glu Met Asp Leu Pro Gly Arg Trp Thr Pro Lys Met Ile Gly 35 40 45	144													
gga att gga ggt ctt gtc aaa gta aga cag tat gat cag ata ccc ata Gly Ile Gly Gly Leu Val Lys Val Arg Gln Tyr Asp Gln Ile Pro Ile 50 55 60	192													

-214-

	ı Ile					Thr					Lev				aca Thr 80	240
Pro	gco Ala	aac Asn	ata 11e	att : Ile : 85	Gly	aga Arg	aat Asn	ctg Leu	ttg Leu 90	Thr	cag Gln	ctt Leu	ggt Gly	tgt Cys 95	act Thr	288
tta Leu	aat Asn	ttt Phe	ccc Pro	Ile	agt Ser	cct Pro	att Ile	gaa Glu 105	Thr	gta Val	cca Pro	gta Val	aaa Lys 110	Leu	aag Lys	336
			Asp					Lys							gaa Glu	384
aaa Lys	ata Ile 130	Lys	gca Ala	tta Leu	gta Val	gaa Glu 135	att Ile	tgt Cys	aca Thr	gaa Glu	ttg Leu 140	gaa Glu	aag Lys	gaa Glu	gga Gly	432
aaa Lys 145	Ile	tca Ser	aaa Lys	att Ile	999 Gly 150	cct Pro	gaa Glu	aat Asn	cca Pro	tac Tyr 155	aat Asn	act Thr	cca Pro	gtg Val	ttt Phe 160	480
gcc Ala	ata Ile	aag Lys	aaa Lys	aaa Lys 165	gac Asp	agt Ser	act Thr	aaa Lys	tgg Trp 170	aga Arg	aaa Lys	tta Leu	gta Val	gat Asp 175	ttc Phe	528
aga Arg	gaa Glu	ctt Leu	aat Asn 180	aag Lys	aga Arg	act Thr	caa Gln	gac Asp 185	ttc Phe	tgg Trp	gaa Glu	gtt Val	caa Gln 190	tta Leu	gga Gly	576
ata Ile	cca Pro	cat His 195	cct Pro	gca Ala	gga Gly	tta Leu	aaa Lys 200	aag Lys	aaa Lys	aaa Lys	tca Ser	gta Val 205	aca Thr	gta Val	ctg Leu	624
gat Asp	gtg Val 210	ggt Gly	gat Asp	gca Ala	tat Tyr	ttt Phe 215	tca Ser	gtt Val	ccc Pro	tta Leu	gac Asp 220	aag Lys	gac Asp	ttt Phe	agg Arg	672
aaa Lys 225	tat Tyr	act Thr	gca Ala	ttt Phe	acc Thr 230	ata Ile	cct Pro	agt Ser	aca Thr	aac Asn 235	aat Asn	gag Glu	aca Thr	cca Pro	999 Gly 240	720
att Ile	aga Arg	tat Tyr	cag Gln	tac Tyr 245	aat Asn	gtg Val	ctt Leu	cca Pro	cag Gln 250	gga Gly	tgg Trp	aaa Lys	gga Gly	tca Ser 255	cca Pro	768
gca Ala	ata Ile	ttc Phe	caa Gln 260	agc Ser	agc Ser	atg Met	Thr	aaa Lys 265	atc Ile	tta Leu	gat Asp	cct Pro	ttt Phe 270	aga Arg	aag Lys	816
caa Gln	aat Asn	cca Pro 275	gac Asp	ata Ile	gtt Val	atc Ile	tgt Cys 280	caa Gln	tac Tyr	atg Met	gat Asp	gat Asp 285	ttg Leu	tat Tyr	gta Val	864
gga Gly	tct Ser 290	gac Asp	tta Leu	gaa Glu	Ile	999 Gly (295	cag Gln	cat His	aga Arg	Thr	aaa Lys 300	ata Ile	gag Glu	gaa Glu	ctg Leu	912

-215-

aga Arg 305	g Glu	cat His	ctg Leu	tgg Trp	aag Lys 310	Trp	ggg	ttt Phe	tac Tyr	aca Thr	Pro	a gad Asp	aaa Lys	aaa Lys	cat His 320	960
Cac Glr	g aaa Lys	gaa Glu	cct Pro	ccg Pro 325	tto Phe	ctc Leu	tgg Trp	atg Met	ggt Gly 330	Tyr	gaa Glu	cto Lev	cat His	cct Pro 335	gat Asp	1008
aaa Lys	tgg Trp	aca Thr	gta Val 340	Gln	cct Pro	ata Ile	gtg Val	ctg Leu 345	Pro	gaa Glu	aaa Lys	gac Asp	agc Ser 350	Trp	act Thr	1056
gto Val	aat Asn	gac Asp 355	Ile	cag Gln	aag Lys	tta Leu	gtg Val 360	gga Gly	aaa Lys	ttg Leu	aac Asn	tgg Trp 365	Ala	agt Ser	cag Gln	1104
	tat Tyr 370															1119
<210> 118 <211> 979 <212> PRT <213> Human Immunodificiency Virus																
	0> 1: Ile		Pro	Ile 5	Glu	Thr	Val	Pro		Lys	Leu	Lys	Pro		Met	
	Gly	Pro	Lys		Lys	Gln	Trp	Pro 25	10 Leu	Thr	Glu	Glu	Lys	15 Ile	Lys	
Ala	Leu	Val	Glu	Ile	Cys	Thr	Glu 40		Glu	Lys	Glu	Gly 45	Lys	Ile	Ser	
Lys	Ile 50		Pro	Glu	Asn	Pro 55		Asn	Thr	Pro	Ile 60		Ala	Ile	Lys	
Lys 65		Asp	Ser	Thr	Lys 70	Trp	Arg	Lys	Leu	Val 75		Phe	Arg	Glu	Leu 80	
Asn	Lys	Arg	Thr	Gln 85	Asp	Phe	Trp	Glu	Val 90		Leu	Gly	Ile	Pro 95		
Pro	Ala	Gly	Leu 100	Lys	Gln	Lys	Lys	Ser 105	Val	Thr	Ile	Leu	Asp 110		Gly	
Asp	Ala	Tyr 115	Phe	Ser	Val	Pro	Leu 120	Asp	Glu	Gly	Phe	Arg 125		Tyr	Thr	
Ala	Phe 130	Thr	Ile	Pro	Ser	Arg 135	Asn	Asn	Glu	Thr	Pro 140	Gly	Ile	Arg	Tyr	
Gln 145	Tyr	Asn	Val		Pro 150	Gln	Gly	Trp		Gly 155		Pro	Ala	Ile	Phe 160	
Gln	Ser	Ser	Met			Ile	Leu	Glu	Pro 170	Phe	Arg	Lys	Gln	Asn 175	Pro	
Glu	Ile	Val	Ile 180	Tyr	Gln	Tyr		Asp 185		Leu	Tyr	Val	Gly 190		Asp	
Leu	Glu	Ile 195		Gln	His	Arg			Ile	Glu	Glu	Leu 205	Arg	Gly	His	
Leu	Leu 210		Trp	Gly	Phe	Thr '		Pro	Asp		Lys 220		Gln	Lys	Glu	
Pro 225		Phe	Leu	Trp	Met 230	Gly	Tyr	Glu	Leu	His 235	Pro	Asp	Lys	Trp	Thr 240	
	Gln	Pro	Ile			Pro (Glu :	Lys			Trp	Thr	Val	Asn 255		

Ile Gln Lys Leu Val Gly Lys Leu Asn Trp Ala Ser Gln Ile Tyr Ala Gly Ile Lys Val Arg Gln Leu Cys Lys Leu Leu Arg Gly Thr Lys Ala Leu Thr Glu Val Ile Pro Leu Thr Glu Glu Ala Glu Leu Glu Leu Ala Glu Asn Arg Glu Ile Leu Lys Glu Pro Val His Gly Val Tyr Tyr Asp Pro Ser Lys Asp Leu Ile Ala Glu Ile Gln Lys Gln Gly Gln Gly Gln Trp Thr Tyr Gln Ile Tyr Gln Glu Pro Phe Lys Asn Leu Lys Thr Gly Lys Tyr Ala Arg Met Arg Gly Ala His Thr Asn Asp Val Lys Gln Leu Thr Glu Ala Val Gln Lys Ile Thr Thr Glu Ser Ile Val Ile Trp Gly Lys Thr Pro Lys Phe Lys Leu Pro Ile Gln Lys Glu Thr Trp Glu Thr Trp Trp Thr Glu Tyr Trp Gln Ala Thr Trp Ile Pro Glu Trp Glu Phe Val Asn Thr Pro Pro Leu Val Lys Leu Trp Tyr Gln Leu Glu Lys Glu Pro Ile Val Gly Ala Glu Thr Phe Tyr Val Asp Gly Ala Ala Asn Arg Glu Thr Lys Leu Gly Lys Ala Gly Tyr Val Thr Asn Arg Gly Arg Gln Lys Val Val Thr Leu Thr Asp Thr Thr Asn Gln Lys Thr Glu Leu Gln Ala Ile Tyr Leu Ala Leu Gln Asp Ser Gly Leu Glu Val Asn Ile Val Thr Asp Ser Gln Tyr Ala Leu Gly Ile Ile Gln Ala Gln Pro Asp Gln Ser Glu Ser Glu Leu Val Asn Gln Ile Ile Glu Gln Leu Ile Lys Lys Glu Lys Val Tyr Leu Ala Trp Val Pro Ala His Lys Gly Ile Gly Ser Pro Ile Glu Thr Val Pro Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn Thr Pro Ile Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Gln Lys Lys Ser Val Thr Ile Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp Glu Gly Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Arg Asn Asn Glu Thr Pro Gly Ile Arg Tyr Gln Tyr Asn Val Leu Pro Gln Gly Trp Lys Gly Ser Pro Ala Ile Phe Gln Ser Ser Met Thr Arg Ile Leu Glu Pro Phe Arg Lys Gln Asn Pro Glu Ile Val Ile Tyr Gln Tyr Met Asp Asp Leu Tyr Val Gly Ser Asp Leu Glu Ile